

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

Yu, Misook

From: Yu, Misook
Sent: Wednesday, September 11, 2002 11:10 AM
To: STIC-Biotech/ChemLib; Chan, Christina
Subject: Request for Rush SEARCH for 09/811,045

This case is due this bi-week.
Please search SEQ ID NO:1 and DNA encoding SEQ ID NO:1.

Chris, please approve the rush search.

Thank you.

Examiner Misook Yu, Ph.D.
703-308-2454 (Phone)
Art Unit 1642
CM1-8E18 (Room)
CM1-8E12 (Mail Box)

*Now teach SEQ ID NO:1 even their own PNAS paper does not
match. 1997 onco gene paper very similar to SEQ ID NO:1
but the date not good.*

THIS PAGE BLANK (USPTO)

STIC-Biotech/ChemLib

75385

Fr m: Chan, Christina
Sent: Wednesday, September 11, 2002 11:25 AM
To: Yu, Misook; STIC-Biotech/ChemLib
Subj ct: RE: Request for Rush SEARCH for 09/811,045

Please rush. Thanks Chris

-----Original Message-----

From: Yu, Misook
Sent: Wednesday, September 11, 2002 11:10 AM
T : STIC-Biotech/ChemLib; Chan, Christina
Subject: Request for Rush SEARCH for 09/811,045

This case is due this bi-week.
Please search SEQ ID NO:1 and DNA encoding SEQ ID NO:1.

Chris, please approve the rush search.

Thank you.

Examiner Misook Yu, Ph.D.
703-308-2454 (Phone)
Art Unit 1642
CM1-8E18 (Room)
CM1-8E12 (Mail Box)

Point of Contact:
Susan Hanley
Technical Info. Specialist
CM1 6B05 Tel: 305-4053

RECEIVED
SEP 11 2002
STIC

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

10-10-1963

THIS PAGE BLANK (USPTO)

09/811,045

WEST Search History

DATE: Thursday, September 12, 2002

<u>Set Name</u> side by side	<u>Query</u>	<u>Hit Count</u>	<u>Set Name</u> result set
---------------------------------	--------------	------------------	-------------------------------

DB=USPT; PLUR=YES; OP=OR

L3	L2 and antibody	3	L3
L2	L1 and hnRNP	3	L2
L1	p2p	55	L1

END OF SEARCH HISTORY

THIS PAGE BLANK (USPTO)

Your SELECT statement is:
s p2p

Items	File
19	5: Biosis Previews(R)_1969-2002/Sep W1
59	34: SciSearch(R) Cited Ref Sci_1990-2002/Sep W2
3	35: Dissertation Abs Online_1861-2002/Aug
26	65: Inside Conferences_1993-2002/Sep W2
4	71: ELSEVIER BIOBASE_1994-2002/Sep W2
11	73: EMBASE_1974-2002/Aug W4
47	94: JICST-EPlus_1985-2002/Jul W2
13	98: General Sci Abs/Full-Text_1984-2002/Aug
60	144: Pascal_1973-2002/Sep W2
2	149: TGG Health&Wellness DB(SM)_1976-2002/Sep W1
12	155: MEDLINE(R)_1966-2002/Sep W2
3	156: ToxFile_1965-2002/Sep W2
3	159: Cancerlit_1975-2002/Jul
1	162: CAB HEALTH_1983-2002/Jul
1	172: EMBASE Alert_2002/Sep W2
1	266: FEDRIP_2002/Jul
2	369: New Scientist_1994-2002/Aug W2
19	399: CA SEARCH(R)_1967-2002/UD=13711
4	434: SciSearch(R) Cited Ref Sci_1974-1989/Dec
1	442: AMA Journals_1982-2002/Aug B1

SYSTEM:OS - DIALOG OneSearch

File 5:Biosis Previews(R) 1969-2002/Sep W1
(c) 2002 BIOSIS

***File 5: Alert feature enhanced for multiple files, duplicates removal, customized scheduling. See HELP ALERT.**

File 35:Dissertation Abs Online 1861-2002/Aug
(c) 2002 ProQuest Info&Learning

File 65:Inside Conferences 1993-2002/Sep W2
(c) 2002 BLDSC all rts. reserv.

File 144:Pascal 1973-2002/Sep W2
(c) 2002 INIST/CNRS

File 155:MEDLINE(R) 1966-2002/Sep W2

***File 155: Alert feature enhanced for multiple files, duplicates removal, customized scheduling. See HELP ALERT.**

File 399:CA SEARCH(R) 1967-2002/UD=13711
(c) 2002 American Chemical Society

***File 399: Use is subject to the terms of your user/customer agreement. Alert feature enhanced for multiple files, etc. See HELP ALERT.**

4/9/1 (Item 1 from file: 5)

DIALOG(R)File 5:Biosis Previews(R)
(c) 2002 BIOSIS. All rts. reserv.

13655648 BIOSIS NO.: 200200284469

P2P -R protein localizes to the nucleolus of interphase cells and the periphery of chromosomes in mitotic cells which show maximum P2P -R immunoreactivity.

AUTHOR: Gao Sizhi; Witte Michael M; Scott Robert E(a)

AUTHOR ADDRESS: (a)University of Tennessee Health Science Center, 3 North Dunlap Street, Room N206, Memphis, TN, 38163**USA E-Mail: rescott@utmem.edu

JOURNAL: Journal of Cellular Physiology 191 (2):p145-154 May, 2002

MEDIUM: print

ISSN: 0021-9541

DOCUMENT TYPE: Article

RECORD TYPE: Abstract

LANGUAGE: English

ABSTRACT: P2P -R is a nuclear protein that can bind both p53 and Rb1. Its functions include roles in the control of RNA metabolism, apoptosis, and p53-dependent transcription. The expression of P2P -R also is repressed in G1 arrested terminally differentiated cells. The current studies therefore evaluated if P2P -R undergoes cell cycle-associated changes in its abundance and/or localization. Western blots show that relative to G0 quiescent cells, P2P -R protein levels are higher in

THIS PAGE BLANK (USPTO)

populations of G2/M cells prepared by the physiological parasynchronization technique of serum deprivation followed by serum stimulation. More striking is the > 10-fold enrichment of **P2P -R** protein in specimens of highly purified mitotic cells prepared by the mitotic shake-select technique, or by synchrony with the mitotic spindle disruption agents nocodazole or vinblastine. These changes in **P2P -R** protein occur without a concomitant change in **P2P -R** mRNA expression suggesting that **P2P -R** immunoreactivity increases during mitosis. Confocal microscopy next established the localization of **P2P -R** to nucleoli in interphase cells and at the periphery of chromosomes in mitotic cells that lack nucleoli. The high levels of **P2P -R** localized to the periphery of chromosomes in mitotic cells suggest that **P2P -R** shares characteristics with other nucleolar proteins that associate with the periphery of chromosomes during mitosis. These include: nucleolin, B23, Ki67, and fibrillarin.

4/9/2 (Item 2 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2002 BIOSIS. All rts. reserv.

13653965 BIOSIS NO.: 200200282786

cDNA encoding P2P proteins and use of P2P cDNA derived antibodies and antisense reagents in determining the proliferative potential of normal, abnormal, and cancer cells in animals and humans.

AUTHOR: Scott Robert E(a)
AUTHOR ADDRESS: (a)Memphis, TN**USA
JOURNAL: Official Gazette of the United States Patent and Trademark Office
Patents 1257 (2):pNo Pagination Apr. 9, 2002
MEDIUM: e-file
PATENT NUMBER: US 6368790 PATENT DATE GRANTED: April 09, 2002 20020409
PATENT ASSIGNEE: University of Tennessee Research Corporation
PATENT COUNTRY: USA
ISSN: 0098-1133
DOCUMENT TYPE: Patent
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: Terminal differentiation is associated with repression in the expression of the **P2P** subset of hnRNP proteins. The present invention provides a **P2P** cDNA that encodes proteins with domains for hnRNP association and Rb1 binding. Probes to the **P2P** cDNA detect a single 8 kb mRNA in multiple murine tissues, in proliferating murine 3T3 cells but not in terminally differentiated 3T3T adipocytes. The interaction of **P2P** cDNA products and Rb1 may serve to modulate cell proliferation and/or other biological functions associated with tumor suppression by an RNA processing mechanism. Moreover, it was shown that **P2P** antisense oligonucleotides selectively repressed 30-40 kDa **P2P** expression.

4/9/3 (Item 3 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)
(c) 2002 BIOSIS. All rts. reserv.

10834146 BIOSIS NO.: 199799455291

The proliferation potential protein-related (P2P -R) gene with domains encoding heterogeneous nuclear ribonucleoprotein association and Rb1 binding shows repressed expression during terminal differentiation.

AUTHOR: Witte Michael M; Scott Robert E
AUTHOR ADDRESS: Dep. Pathol., Univ. Tennessee Med. Cent., Memphis, TN 38163**USA
JOURNAL: Proceedings of the National Academy of Sciences of the United States of America 94 (4):p1212-1217 1997
ISSN: 0027-8424
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: Terminal differentiation is associated with repression in the expression of the proliferation potential proteins (**P2P**) subset of heterogeneous nuclear ribonucleoprotein (hnRNP) proteins. We report here the cloning and characterization of a 5173-bp **P2P** -related (**P2P -R**) cDNA that contains a 4214-bp open reading frame. Probes to this cDNA detect a single 8-kb mRNA in multiple murine tissues and in proliferating

THIS PAGE BLANK (USPTO)

3T3T cells, but not in terminally differentiated 3T3T adipocytes. Evidence that this cDNA can encode peptides with domains for hnRNP association was established by showing that such peptides are recognized by two monoclonal antibodies known to detect core hnRNP proteins, and by showing that the C130 monoclonal antibody, produced against a cDNA-derived fusion protein, also selectively detects native **P2P** hnRNP proteins. In addition, **P2P** -R cDNA-derived fusion proteins bind single-stranded nucleic acids, and a **P2P** -R cDNA-derived antisense oligonucleotide selectively represses **P2P** expression. Because terminal differentiation is associated with modulation in Rb1 function, we assayed if products of this cDNA might interact with Rb1. Evidence that the **P2P** -R cDNA encodes a protein domain that binds Rb1 was established using a glutathione S-transferase fusion protein to selectively precipitate Rb1 from cellular extracts. Data also show that this binding is reduced by competition with the adenovirus Eta protein, indicating that binding occurs through the "pocket" domain of Rb1. These results establish that the **P2P** -R cDNA encodes protein domains involved in both hnRNP association and Rb1 binding and complement recent reports that localize Rb1 to sites of **RNA** processing in the nucleus.

4/9/5 (Item 1 from file: 399)
DIALOG(R) File 399:CA SEARCH(R)
(c) 2002 American Chemical Society. All rts. reserv.

136261811 CA: 136(17)261811a PATENT
P2P protein, cDNA, antibodies and antisense oligonucleotides for determination of proliferative potential of normal, abnormal, and cancer cells in animals and humans
INVENTOR(AUTHOR): Scott, Robert E.
LOCATION: USA
ASSIGNEE: The University of Tennessee Research Corporation
PATENT: U.S. Pat. Appl. Publ. ; US 20020035080 A1 DATE: 20020321
APPLICATION: US 811045 (20010316) *US PV27568 (19960927) *US 801308 (19970218)
PAGES: 32 pp., Division of U. S. Ser. No. 801,308. CODEN: USXXCO
LANGUAGE: English CLASS: 514044000; A61K-048/00A; A61K-038/17B;
C07K-014/435B; C07H-021/02B; C07H-021/04B
SECTION:
CA215002 Immunochemistry
CA203XXX Biochemical Genetics
CA209XXX Biochemical Methods
CAS REGISTRY NUMBERS:
405051-29-8P amino acid sequence; P2P proliferation potential proteins, antibodies, nucleic acids and antisense oligonucleotides for diagnosis and treatment of proliferative diseases
405051-30-1P nucleotide sequence; P2P proliferation potential proteins, antibodies, nucleic acids and antisense oligonucleotides for diagnosis and treatment of proliferative diseases
405006-95-3 P2P proliferation potential proteins, antibodies, nucleic acids and antisense oligonucleotides for diagnosis and treatment of proliferative diseases
405051-74-3 unclaimed nucleotide sequence; p2P protein, cDNA, antibodies and antisense oligonucleotides for detn. of proliferative potential of normal, abnormal, and cancer cells in animals and humans

4/9/6 (Item 2 from file: 399)
DIALOG(R) File 399:CA SEARCH(R)
(c) 2002 American Chemical Society. All rts. reserv.

132304218 CA: 132(23)304218z JOURNAL
Gene-target recognition among members of the Myc superfamily and implications for oncogenesis
AUTHOR(S): O'Hagan, Ronan C.; Schreiber-Agus, Nicole; Chen, Ken; David, Gregory; Engelman, Jeffrey A.; Schwab, Richard; Alland, Leila; Thomson, Cole; Ronning, Donald R.; Sacchettini, James C.; Meltzer, Paul; DePinho, Ronald A.
LOCATION: Department of Adult Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA
JOURNAL: Nat. Genet. DATE: 2000 VOLUME: 24 NUMBER: 2 PAGES: 113-119
CODEN: NGENEC ISSN: 1061-4036 LANGUAGE: English PUBLISHER: Nature

THIS PAGE BLANK (USPTO)

CAS REGISTRY NUMBERS:

74812-49-0 gene Cull1 subunit Cullin 1; gene-target recognition among members of Myc superfamily and implications for oncogenesis - common targets for Myc and Myc(Bx11-basic region) chimeric protein
9024-60-6 gene ODC1; gene-target recognition among members of Myc superfamily and implications for oncogenesis - common targets for Myc and Myc(Bx11-basic region) chimeric protein
9013-05-2 gene P2R4; gene-target recognition among members of Myc superfamily and implications for oncogenesis - Myc-specific targets
9026-43-1 gene RING3; gene-target recognition among members of Myc superfamily and implications for oncogenesis - Myc(Bx11-basic region) chimeric protein-specific targets
120178-12-3 gene Tert; gene-target recognition among members of Myc superfamily and implications for oncogenesis - common targets for Myc and Myc(Bx11-basic region) chimeric protein
80449-01-0 gene TOP1; gene-target recognition among members of Myc superfamily and implications for oncogenesis - Myc-specific targets
74870-74-9 gene UMPS; gene-target recognition among members of Myc superfamily and implications for oncogenesis - Myc-specific targets
4/9/7 (Item 3 from file: 399)
DIALOG(R) File 399:CA SEARCH(R)
(c) 2002 American Chemical Society. All rts. reserv.

111229734 CA: 111(25)229734a JOURNAL

Loss of proliferative potential during terminal differentiation coincides with the decreased abundance of a subset of heterogeneous ribonuclear proteins

AUTHOR(S): Minoo, Parviz; Sullivan, William; Solomon, Larry R.; Martin, Terence E.; Toft, David O.; Scott, Robert E.

LOCATION: Dep. Biochem. Mol. Biol., Mayo Clin./Found., Rochester, MN, 55905, USA

JOURNAL: J. Cell Biol. DATE: 1989 VOLUME: 109 NUMBER: 5 PAGES: 1937-46 CODEN: JCLBA3 ISSN: 0021-9525 LANGUAGE: English

SECTION:

CA213006 Mammalian Biochemistry

IDENTIFIERS: ribonucleoprotein cell differentiation, proliferation potential protein cell differentiation

DESCRIPTORS:

Proteins, specific or class, P2P (proliferation potential protein)...

Ribonucleoproteins, heterogeneous nuclear RNA-contg....

of cell of human and lab. animal in terminal differentiation

Mesenchyme, stem cell...

ribonucleoproteins of, in terminal differentiation

Cell nucleus...

ribonucleoproteins of, of cells of human and lab. animal in terminal differentiation

Skin, keratinocyte, composition...

ribonucleoproteins of, of human in terminal differentiation

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2002, 15:21:20 ; Search time 63.16 Seconds

(without alignments)
2469.088 Million cell updates/sec

Title: US-09-811-045a-1
Perfect score: 7374
Sequence: 1 MAEKVDPNNKGAMLTNTGKY.....WRRSWLGRWRAPSSRNPS 1404

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_032802.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4922	66.7	1278	22	AAAM39259
2	3497	47.4	973	22	AAAM1045
3	560.5	7.6	155	22	AAAG3720
4	542	7.4	1231	22	ABBS9738
5	531	7.2	146	22	AAO02604
6	446	6.0	153	21	AAAT0234
7	352	4.8	3910	14	AAAR38470
8	352	4.8	3910	16	AAAR6462
9	351	4.8	2768	22	ABBB68397
10	344.5	4.7	3111	22	ABBB0327
11	340	4.6	2703	22	ABBB0074

12	340	4.6	3660	22	ABG12812	Novel human diago
13	335.5	4.5	1026	22	AAAM7825	Human protein seq
14	333.5	4.5	2519	22	ABG16636	Novel human diago
15	332.5	4.5	1151	22	ABBB1598	Drosophila melanog
16	331	4.5	2897	22	ABBB5814	Drosophila melanog
17	330	4.5	2951	22	ABBB60291	Drosophila melanog
18	330	4.5	1430	22	ABBB58724	Drosophila melanog
19	329.5	4.5	617	22	ABBB5445	Peptide #2951 enco
20	329.5	4.5	617	22	ABBB20884	Protein #2883 enco
21	329.5	4.5	617	22	AAAM6269	Human brain expres
22	329.5	4.5	617	22	AAAM16458	Peptide #2892 enco
23	329.5	4.5	617	22	AAAM04187	Peptide #2869 enco
24	328	4.4	1033	22	AAAM79809	Human protein seq
25	328	4.4	2819	22	AAAB35408	Human 07CG27 gene
26	326.5	4.4	1341	21	AAAB5657	Human Aclous L pro
27	326.5	4.4	1342	22	ABBB09943	Novel human diago
28	324.5	4.4	2663	22	AAAB98612	Human tumour suppr
29	324	4.4	2724	22	ABBB20119	Novel human diago
30	323.5	4.4	1333	22	ABBB66754	Drosophila melanog
31	323	4.4	1524	22	ABBB04845	Novel human diago
32	321	4.4	3969	15	AAAB52971	Product of the cDN
33	318	4.3	1163	22	AAU28028	Novel human secret
34	318	4.3	5533	22	ABBB65772	Drosophila melanog
35	318	4.3	5560	22	ABBB71160	Drosophila melanog
36	315.5	4.3	3080	22	ABBB4877	Drosophila melanog
37	309	4.2	3257	22	ABBB7502	Drosophila melanog
38	309	4.2	6815	22	ABBB6811	Drosophila melanog
39	306.5	4.2	955	21	AAAB4231	Human ORFX ORF195
40	306.5	4.2	962	22	ABBB11976	Human TRAP150 homo
41	304.5	4.1	1183	22	ABBB58769	Drosophila melanog
42	304	4.1	931	22	ABBB1093	Drosophila melanog
43	303	4.1	724	21	AAAG46505	Arabidopsis thalia
44	303	4.1	778	21	AAAG46504	Arabidopsis thalia
45	301.5	4.1	1984	22	ABBB1060	Drosophila melanog

ALIGNMENTS

RESULT 1	
AAAM39259	standard; Protein: 1278 AA.
XX	
XX	AAAM39259;
AC	
XX	
DT	22-OCT-2001 (first entry)
XX	
XX	
DE	Human polypeptide seq ID NO 2404.
XX	
KW	Human: nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW	leukemia.
XX	
OS	Homo sapiens.
XX	
PN	WO20015312-A1.
XX	
PD	26-JUL-2001.
XX	
PR	26-DEC-2000; 2000WO-US34263.
XX	
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0598042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
XX	

PA (HYSE-) HYSEO INC.
XX Tang YR, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
DR WPI: 2001-44253/47.
DR N-PSDB: AA158415.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PS Example 4; SEQ ID NO 2404; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA158642-AA162213) with nootropic,
CC immunosuppressant and cytosstatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 1278 AA:
SQ

Query Match 66.7%; Score 4922; DB 22; Length 1278;
Best Local Similarity 75.2%; Pred. No. 1.5e-286;
Matches 972; Conservative 81; Mismatches 148; Indels 92; Gaps 15;

QY 1 MMEVKDPNNKGAALMTWTKAIPPTIDAEVAIGKKKKPPPLPEEPSSSEEDDPIFAELL 60
DB 23 mmevkdpnnkgaalmtwtgfaipptidaeaaygkkekppllpeepssseeddpiidell 82
QY 61 CLICKDINTDAVVIPCCGNSSCDECIKRTTLLESDEKHTCPTCHONDVSPDALANKFLRQA 120
DB 83 clickdimtdavvipccgnssycdecirtaallesdehtcpcqndvspdallankflrq 142
QY 121 VNNFKETGTYTRLRKQOLPEFLFVPPRPRLSQRNLOQRRSRITLNOQDQVVRVRYVSPR 180
DB 143 vnnfketytkrlrkqolpeflfvprrpplrqnlqrlmrspisrqqdplmiprvtsst 202
QY 181 CSDTKTAGSCSDSGTSLRPLPAPISISLTSNQSILAPPVSGNPSAPAPVDIATVTSISY 240
DB 203 -----hpapsissltngssilappvsgnpsapapvdiaatvtsisy 244
QY 241 HSEKSDGPRDSNKLPLPAALITSEHSKAGSIAITALEEKG---VPGTSPWNSIFWG 296
DB 245 hseksdgpdrdsnkllpaaalitsehskgassiataltaleekgyvpyvgtpse---llg 300
QY 297 QSLTHQQLPTTGPRVAINARPGGGRPGMEHSKGLGVLSPPQOIKRGEKSCYRSINRGR 356
DB 301 qslthqqlpttgprvaintarpgggrpgmehsnklylvsppqdlrrgescyrsinrg 360
QY 357 HHSERQSGTSPSLPAPPCFVPPPPPLVPPPHLPLPPGVPPPOFSPQFSSOPTG 416
DB 361 hhsersqgtspslpactpfvppppplvppphlplppgvpppfspqfppgqpatag 420
QY 417 YSVPPGFPAPANISTACFSPGVPAHNSNTMTQOAPLISREEFYREQ-----465
DB 421 ysvppgfpapanistacfspgvpahnsntmtqoaplsreefyreqrllkeekks 480
QY 466 -----NDKGR-----SKPPYSGSSTSRSSYTDS-SQGLAQHIAHLT 501
DB 481 kldetndfakelmeykklqkerrtsfsrskapsygsysrslvysksrsgsltrrsys 540

QY 502 LSPSAAHITLIDLHDPHP-----PEEAARSAMIVHMPDLMDIAHARSRPYRRRSR 557
DB 541 rfsrshsrsysrpppyrrgrgkarnysrsh-----gynrsrsrpppyrrrsr 595
QY 558 SPPEFRGQSFTKRNVPREEKREYNRYREVPPTYIKAYGASVDPDRDFEERYRWE 617
DB 596 spqafrgspnkrnvpqgelereyfnryrevpppydmkaygrsvdfdrpfeekeryewe 655
QY 618 RKRYREYKRYKGYAAGAPRPSANREDSPRELLPLNRSNFTGRRRDYAAAGSHRW 677
DB 656 rkyreyekyykygaagaprpasaneisrpfllnlnspfltrgyredyggshrs 715
QY 678 RNLGNYPEKLTSTRDSHNAKNPKSKKESEENYPCGDKGNKHKHRRRNEKESESE 737
DB 716 rnlgnypelkltstrdsnhnknpskkskksenapggdkgnkhkhnlrr---kgeesegf 772
QY 738 LNPELLFETSKKCGSSGIDETKTDTLFVLPSRQDATPVYRDEPDASTTFKYSDDKRE 797
DB 773 lnPELLfetskrspgyveenktdslfvlpsridatpvrdpmdaeslftksvsekdkte 832
QY 798 KDKPKVSDTKRKSDGSATAKKDNVLKPSKGPQEKYVDGDRKSPSEPLKAKKEEATK 857
DB 833 rdkpkakgdktkrkdngsavskkenlvkpkqpkvdygeretspspeplkakeetpk 892
QY 858 IDSVKPSSSQKDEKYTGPRKASKSADTTRQSPRRRSKRYVKTSSOKSPVRRR 917
DB 893 tdtvksessqkdekltgprkshksakehgetkpvkeekvkdysk--dvksekltk 950
QY 918 RPRSLRKINYLAREKN-----EREKKR---KSYDKPESSSMKISKVEGEIYKPPS 967
DB 951 eeka-kkpn-----eknkpldnkgkrkkteekgyvkdfeasssmisklvelelvkpsp 1004
QY 968 KRKMGDVEKLEKTPKPKDIASSTPAKKIKILNREGRKIGNAENASTKREPSEKLEST 1027
DB 1005 krkmgdtekmtdrpekdki-slsapakiklnregrkigsenlnkpekelest 1063
QY 1028 SKIKQEKVKAKARKVAGSESSSTLVDTYSTSGSPYRKSEKTDKRYVIKTMEEY 1087
DB 1064 skvqgkvkgykrvtkytgiegssstlvdststsgsprkseeektdkrtvltkmeey 1123
QY 1088 NNDNTAPAEVYIMIOVPOSKWDKDFESEEDVKTQTPQISYGKSSIIKNTTTPSAT 1147
DB 1124 nndntapadvilmiqvpqskwkddeeseedvktqtpqissvgkpasvllknvsttkpsnl 1183
QY 1148 AKYTEKESPOPEKLOKLPKASHELMOHLELRSSKGSASSKGRKAPREHSGSKNDPDR 1207
DB 1184 vkypkesepsekigkftkdvshelilqhevksksnassekgktdardysvlekenpekr 1243
QY 1208 KSGAQPDKESYVDRLSEOGHFKTLQSOSKRETRT 1240
DB 1244 knstgpekesnldrlnegnfkalsgskreart 1276

RESULT 2
AA141045
ID AA141045 standard; Protein; 973 AA.
XX
XX
XX AA141045;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 5976.
XX
XX Human; nootropic; immunosuppressant; cytosstatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX
XX Homo sapiens.
XX
XX

PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0486725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0596042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HXSE-) HXSEQ INC.
 XX
 PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao Qa, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB: AA160201.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 2; SEQ ID NO 5976; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 XX Sequence 973 AA:
 SQ

Query Match 47.4%; Score 3497; DB 22; Length 973;
 Best Local Similarity 72.9%; Pred No. 1.2e-202;
 Matches 693; Conservative 43; Mismatches 121; Indels 94; Gaps 11;

QY 2 MEVKDPNMKGAMLTNTGKAYIPTIDAAYAIQKKEKPPLEPSSSSSEEDPIPAELLIC 61
 DB 48 mevkdpnmkgamltntgkayipidaeyaiqkkekpflepssssseeddipelllc 107
 QY 62 LICKDITMTAVYIPCCGNSCDECIITLTLESKHTCPICQNDVSPDALIANKFTROAV 121
 DB 108 lickdmtavypccgnsycdecirtallesdehtcpcchqdvspdaliankftirgav 167
 QY 122 NMFKNETGYTKRLRKOLFPLVPPRPLSOBNLOPRSRSPILRQODPVVFYFYTSPIC 181
 DB 168 nmfknetytkrlrkolfplvpprpplsobnloprsrspilrqodpvmvfyytsspc 226
 QY 182 SDTKTAGSCSDSGTLRLPAPSISSLTNSQSSLPAPVSGNSAPAPVDITATVSIHV 241
 DB 227 -----hpsisistngslapvsgnpsapapvditativsihv 269
 QY 242 SEKSDGPFKDSNKKLLPAALITSEHSKASITITAMEKG-----VPGTSPMNSTIFVGO 297
 DB 270 seksgdpfkdsnkkllpaalitsehskasitaitameekgyqvvdltfpts---1199 325

QY 298 SLIHGOLIPDTGPVRINARPGSGRPGWEHSNKLGLVSPPOOIRRGERSCYRINRGH 357
 DB 326 slhggolipdtgpvrinarp9ggrpgweshnlgylvspppqitrgescyrlnrgth 385
 QY 358 HSEBSQRTQSPSLPATPCFVWPVPPPLYPPTTLPLPGVPPQFSPQFPSSQPPFAGY 417
 DB 386 hsebsqrtqspslpatpcfvwpvppplypppttlplpgvppqfspqfpppppaa 445
 QY 418 SVPPGFPAPANISTACFSPGVTAHSNMTPTQAPLRSREFFYREQ----- 465
 DB 446 svppgfpapanistacfspgvtahsnmtptqaplsreeffyrekeekkak 505
 QY 466 -----NDKGRE-----SKFPYSSSSYSSSYTDS-SQGLAQHIALTL 502
 DB 506 ldefndfakelmeykikqerrtsfsrskspysssysrskysrsgtsrstrsyr 565
 QY 503 SPSAAHTLDDLHDHPH-----PEEAARSAMIVHMPDLMDIAHARSRSPRYRRSRSS 558
 DB 566 sfershrstsrspyprrirgsksrnyrsrsh-----ghrsrstrpprryhrstrs 620
 QY 559 PPEFGQSPTKRNVPREEKREYFNRYREVPPYDIKAYGRSVDRPPEKEREYEMER 618
 DB 621 pgefgrspknrvpreekreyfnryrevppydikaygrsvdrppekeryewer 680
 QY 619 KYREWETKYYKGAAGAOPRPSANREDPSERLLPLININSPTTRRRRDYAAQSHNR 678
 DB 681 kyrewetykkygaagaprpnsanrenfserflplnlnspfltrdyv99gshsr 740
 QY 679 NLGNPYEKISTPDSHNAADNPKSEKESENVGDKGNKKHRRRREKEESESEFL 738
 DB 741 nlgnypeklsatdgmgndntkskesenapdgkgnkhkhrtrr---ygeesegfl 797
 QY 739 NPELETSKRKCGSSGIDETKDTLVLDSRDATPVDRPDMAESITKVSVDKKRREX 798
 DB 798 npeletskrskcgssgidetkdtlflvsrdatpvrdepmaesitfkvsvekdrer 857
 QY 799 DKRYVSDTKRRSDGSAVAKKDNVLPKSGQPEKVDGDEKSPREBPLKAKKEZATRI 858
 DB 858 dkryvsgdtkrrsdgsavakknvlpksgqpekvdgdevrldlndlqkpkkeetpkd 917
 QY 859 DSV-----KPSSSQKDEKVTGTPRKRAHSKAADTRROQSPRTRRSKR 901
 DB 918 ltlhnhlplrtmkkslepkekt-----lqgktrpktsqr 955

RESULT 3
 AAG73720
 ID AAG73720 standard; Protein; 155 AA.
 XX
 AC AAG73720;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen protein SEQ ID NO:4484.
 XX
 KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma; chromosome 16.
 XX
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26524.
 XX
 PR 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Barash SC, Blirze CE, Rosen CA;
 XX

Db	551	as1ygyvaakvgsppidddpleafirlnkekerkkvdftrssdtrhsrpsd---rqrhrfk	567
Qy	679	NLGNGNP--EKLSTRDSHNAKD-NPKSEKESENVPGDGGKNKHKK-RKRNRDEKGEES	734
Db	588	s-----pnyeksdndr--nlkdkrprsrerkrh-----syerhlhrprsrqndgsks	635
Qy	735	ESFLNPELLETSRKRCRSGSGGIDETKTDTLFLVLPBRDATPYRDEPMMAESITRKSVSKD	794
Db	636	pg-----grikfr--sg-----hrrisasp-----	651
Qy	795	KREKDKPVKSDNTRKRKSDGSATAKNDNLKPSKG--POEKVDGDRKSPRSSEPLLKAK	852
Db	652	-----kpyksdydkpyknkpsapkteavepppgfepqltd---edgyrnknprusssea	703
Qy	853	EEATKIDSVPKSSSSOKDEKVTGTPRRAN-----SKSADKTRROSQPRTTRSKRTVPK-T	906
Db	704	sgsqsgsdsskkrigenrhee---aprktrhsrstskbpknndsnysrltppakltppkmt	759
Qy	907	SSQSKQPVTRTRPSLKRT--NYLIARBK-----NEREKRRKSYDKOPRESSSMKISK	956
Db	760	aaglrq--resspkrtpkeshdylitakarimasgpyndte-metnygkenkaks-plsk	815
Qy	957	VEGTETVAPSPRKMEGVLEKLEPTPEKDKLASSTTPAKTKILNREIGKTIQNAENKSTT	1016
Db	816	-----drkkkkdkdkkaer-----knkkdkktrakekgdqrqkxss	851
Qy	1017	KEPSEKLESTSSKIKOEKVKKAKRKVAGSESSSTLVDTSTSGSPVRKSEKERTDT	1076
Db	852	vnrdd-----sdlnnslnmesnykvlspraqgspl-----elnaaglsplnhatenvp	901
Qy	1077	KRTVTKITEEYNNNDTAPAEVYITIMIOVPOSKWPKDPESEEDPYKTTQPIQSYGKPSST	1136
Db	902	kshsl1itvgaasddnlgpraklseansvnlskwiden1lgled--sskkaagaasddpsel	960
Qy	1137	IKNVTTPKSAFAKYTEKSEQEPKLOKLPRKASHELQHELRSKGSASSPKGRKAKREH	1196
Db	961	tsdvtirkenaalfakalnairpmeftyv-----lnsknksdtrsv	1000
Qy	1197	SGSEKND--NPDKRKSGAQPDKESTVDRL-----SEOGHFETLSQSSKETRTSEKHSVR	1248
Db	1001	vrsddrassprtnms-----stsvkdtlglkklndrstrsrdekekyrrraarsdddanr	1055
Qy	1249	GSSKNDFTPGDKKVDVDSRYS--SKRRDRGELARKRDSPPRGKSLSGQSKLLEER	1307
Db	1056	grsdthgsrkrkdnr---sldrdaapsekrqer---syktrsp-----edoklrrqn	1099
Qy	1308	DLPRKGAASK-----KSNSSPRDKKPHDHKAPYETK-----RPOCELRP-----	1347
Db	1100	---kegssskngkhqdnmsddtrraakntkssdrvssvstlavappkpcdrpdnprtkf	1156
Qy	1348	VDKNSGKER-----EKHAAEARNGKE	1368
Db	1157	vdtsessslvkvkntdlitqkegassdngme	1185
RESULT 5			
ID AAO02604			
AAO02604 standard; Protein: 146 AA.			
XX AAO02604;			
AC XX			
XX XX			
DT 06-NOV-2001 (first entry)			
XX XX			
DE Human polypeptide SEQ ID NO 16496.			
XX XX			
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;			
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;			
KW tissue growth factor; immunomodulatory; cancer; leukaemia;			
KW nervous system disorders; arthritis; inflammation.			
OS Homo sapiens.			
PN MO200164835-A2.			

```

XX 07-SEP-2001.
PD
XX
XX 26-FEB-2001; 2001WO-US04927.
PE
XX
XX 28-FEB-2000; 2000US-0515126.
PR
XX 18-MAY-2000; 2000US-0577409.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YF, Liu C, Drmanac RT;
PI
XX WPI: 2001-514838/56.
DR N-PSDB; AA182535.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PI disorders -
XX
XX Claim 20; SEQ ID NO 16496; 1399pp + Sequence Listing; English.
PS
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
SQ Sequence 146 AA;

```


CC of the putative fingers. The structure of these cysteine-rich domains
 CC appears to be unique to the trithorax and ALL-1 genes.

XX Sequence 3910 AA;

Query Match 4.8%; Score 352; DB 14; Length 3910;
 Best Local Similarity 20.1%; Pred. No. 6e-12;
 Matches 288; Conservative 154; Mismatches 467; Indels 524; Gaps 61;

```

QY 90 LIESDKHTCPTCHQNDVSPDALIANKELRQAVNNENKETYKRLKQLPF----- 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 298 lpspskrt-----dataklqlqra-----kkgagkklekaeqigyrkvcq 340
QY 142 -----LFLVPPRPRLSQRLQPRSRPILRQDDPV-----FRTVSPTCSDTMT 186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 341 vknltqqlmpvvasalsrltkprtlfededydprikarlslastpnrls-apsqsssek 399
QY 187 AGSCSD-----SGTLRLPAPSISSLTNSQS-----LAPVSGNPSAPAPVDITATV 236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 400 ssaasqhsqmsdsrsspsvdtstsdgaeeiqvlpceersdtpvhp-rlp----- 453
QY 237 SISVHSEKDGPPROSDNKLPAALITSEHSKGASIAITALEMEKGVPTSPWNSIEVG 296
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 454 sspenesndrstrys-----vsersfgsrt----- 480
QY 297 QSLHGOQLPTTGPRINARPGGRRGWEHSKLGVLVSPPOQIRGRGSCYRSINRGR 356
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 481 -----LKKLSLsgsqp----- 492
QY 357 HHSEKQRTQSPSLPATPCFVPPPLYPPLPHTLPLPG-----VPP--POFS 404
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 493 -----getsss-----pppplltppp--plpbaasidhtpwlmpptlpas 532
QY 405 POPSSQPPTAGS---VPPPGF-----PPAPANISTACTS-PCV--PTAHSNTMPT 450
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 533 pflpsastapmgkrlksalreptflrwlslksrsepyfssakakeglrlrkpfdnfrpp 592
QY 451 TQAPLISREFREONDKRESKFPYSGSSYSTSDSOGIAOHIIHALITLSPSAHNL 510
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 593 pltp-----edvgfaasqfssglaasarlfsphsgtrfthm--ktspp----- 633
QY 511 DLLDHHPPEEAEARSAMIVHMDLMDIAHARSPPRYRYSRSPREFRQSPTRK 570
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 634 -llraptrfpseahsrifsvtlpsnrttaagussgvsnrkkrkxfspirseprips-- 690
QY 571 NVPREKEKEREYRNRREVPPYDIKAYGRSYD-----FRDPFEKERYREWERKY 620
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 691 hamtrtsgrlsseslaptlpssvssslsivspatlalsnplftfss----- 738
QY 621 REMYEKYVGYAVGAOPR-----SANREDP---SPERLLPLNIRNSPFTGRREDYAAGO 673
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 739 ---hslltqgsaeaknqirpkrqtspaeapfssssplrlfpwfrpsqtegrnkdkaa--- 792
QY 674 SHRNRLNGNVPBEKLS-TRDSHNAKDNPKSKESKESNVGDGCKHHKRRKRNKEKGE 732
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 793 -----peelsktdadaksveksreidrer-----ekenrresrkekrkkysel 837
QY 733 ESESFLNP-----ELLETSRRCGSSGIDETKTDTLFLVLRSDAPVPRDEPMD 781
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 838 gssasalpygrvskkvgedvaltsasakktgirkks-----shdsqg-----d 882
QY 782 ABSITFKSVSDKDKREKDKPKVSKDKTKRKSDGSATAKDNVL-KRSKGQEKVVDGDRK 840
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 883 ltsvlt-----gdtavtkllikrgynlektnd-1 914
QY 841 SPRSEPLKAKE---EATKIDSVKPSSSS-----OKDKVGTCPKRAKSKSA 885
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 915 gp-tapslektektlclstpsstsvkhstssigmldgacklpmtdkrrvasllkkaqaqlc 973
QY 886 KQTRROSQPTRRSRKRTVPKTSQSKQPVRT--RRPRSLAKINLYLAREKNEREKRRKSV 943
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 974 klekskslkqtdq-----pkagqgesdssetsvrgpr-----lkh-vorraavaigrkrav 1023

```

```

QY 944 DKDFESSMKISVEGETELVKKPSKKMEGDYEKLERTP--EKDKTASSTTPAKKIKILNR 1001
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1024 fdp-----dmptlsalrpeeteklss-----ngn 1048
QY 1002 ETGKKIGNAENASTPEPEKLES-TSSKIKQEK--VKGKAKRKVAGSBC-----SSSTL 1053
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1049 ddksssiagsdeaplrlpklrvtrnkapepprkkgrrrrrcgpcgpcgpcdcgyc 1108
QY 1054 VDTSTSTSGSFPVRSKSEKTDTKTVIKTMEBYNNNDNTAPADVIIMIQVPOSKWDKD 1113
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1109 tncldkprktfgynlkkqck-----mrkcgqlqympskaylqkqakavkkkexks 1158
QY 1114 FESEEDVATTQPIQSVGPPSIITKNVTTKPSATAVYTEKESRQPEKQLKLPREASHHELM 1173
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1159 ksekkd-----skessvkvknv-----dsqkp----- 1182
QY 1174 QHELRSKGSASSEKGRANDREHSGSEKDNPDKRSKGAOPDKESVTRLSEOGHFETLSQ 1233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1183 -----tparedpapkksessepprkpyveekseegnvasapp 1219
QY 1234 SSKETRTSEKHESVRCSSMKDF-----TPGRDKRVYDSDRYSSSKRRDERGELARRK 1286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1220 eskqatlpasrksskqvsqpalvlpqpptltgprk-----evpkl 1261
QY 1287 DSPPRGESLSGQSKTLREERDLPKKGAESKSNSSPPRDKRPHDKAPETRRPCEETK 1346
   : : : : : : : : : : : : | : | : | : | : | : | : | : | : | : |
Db 1262 psepkkkqp-----pppesqpegskqkqvaprlpvlvkqk--pkekxp---p 1304
QY 1347 PVDK-----NSGKER-----EKHAAEARNRGRESSG 1371
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1305 pvnkgqagntlnlslngnsqkqlpadgyhrrlvdfkdeceaanvwmng 1357

```

RESULT 8
 AAR66462
 ID AAR66462 standard; Protein; 3910 AA.
 XX
 AC AAR66462;
 DT 22-AUG-1995 (first entry)
 XX
 DE ALL-1 (acute lymphocytic leukaemia-1) protein.
 XX
 KW Acute lymphoblastic leukaemia; acute nonlymphoblastic leukaemia;
 KW chromosomal translocation; abnormality; detection; t(4;11);
 KW t(9;11); t(11;17); ALL-1; AF-4; AF-9; AF-17; chimeric gene; probe.
 XX
 OS Homo sapiens.
 FH
 FH Key Location/Qualifiers
 FT Protein 348..3910
 FT /note= "in the disclosure, numbering of amino acid
 FT residues starts from Met348 as position 1"
 FT
 FT Region 1369..1569
 FT /note= "cysteine-rich region of homology to
 FT Drosophila trithorax; conty. sequence
 FT motifs analogous to zinc finger domains"
 FT
 FT Domain 1372..1397
 FT /label= "zinc finger
 FT /note= "homologous to zinc finger domain 3 of
 FT Drosophila trithorax"
 FT
 FT Domain 1417..1447
 FT /label= "zinc_finger
 FT /note= "homologous to zinc finger domain 4 of
 FT Drosophila trithorax"
 FT
 FT Domain 1465..1497
 FT /label= "zinc_finger
 FT /note= "homologous to zinc finger domain 5 of
 FT Drosophila trithorax"
 FT
 FT Domain 1507..1537
 FT /label= "zinc_finger
 FT /note= "homologous to zinc finger domain 6 of

Db 2308 v--spvdfdvpssekpvav-----seydgeesteppvhdvetstdepts----- 2349

QY 1120 DVKTQPIQISGKRS-----SIINKNTTKPSATAKYTEKSESEPEKIQIKPKASHRLM 1173

Db 2350 daklkrptseapateapatealvpeaapelekevepeakepelekepeake--- 2406

QY 1174 QHELRSSKGSASSEKGRAKDREHSGSEKDNPKRRSGAQDPKESTVDRLSFGHFKTLSQ 1233

Db 2407 qpel-----eketpkeatepeleke-tpkeateq-----pe 2437

QY 1234 SSKRT--RTSEKHPSVKGSSKKDFTPGRRKKVVDYSDRYSSSKRRDERGELARRKDSPPR 1291

Db 2438 lkevtlkatkeapesv-----dektlpepvvkpslstdsteedeeseesee--esaadkdd--k 2489

QY 1292 GKESLSQGSKLTREERDLPKKGAE--SKSNSPPRDKKP--HDHKAPYETKRRCPEETKPYD 1349

Db 2490 nket-eedtdkhhneepavvsei pqseeaavptqthlptlhassttppa-----vd 2543

QY 1350 KNSGKEREKH 1359

Db 2544 drygeedeen 2553

RESULT 10

ABB60327 standard; Protein; 3111 AA.

XX ABB60327;

AC ABB60327;

XX 26-MAR-2002 (first entry)

DT 26-MAR-2002

XX Drosophila melanogaster polypeptide SEQ ID NO 7773.

DE Drosophila melanogaster polypeptide SEQ ID NO 7773.

XX Drosophila; developmental biology; cell signalling; insecticide;

KM pharmaceutical.

KW pharmaceutical.

XX Drosophila melanogaster.

OS Drosophila melanogaster.

XX WC200171042-A2.

PN 27-SEP-2001.

PD 23-MAR-2001; 2001MO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PA Venter JC, Adams M, Li PWD, Myers EW;

PI WPI; 2001-656860/75.

DR N-PSDB; ABL04430.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

PS Disclosure; SEQ ID NO 7773; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

CC sequences (AB101840-AB16175) and the encoded proteins

CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pcl_sequences.

XX

SO Sequence 3111 AA;

Query Match 4.78; Score 344.5; DB 22; Length 3111;

Best Local Similarity 19.18; Pred. No. 1.3e-11;

Matches 264; Conservative 172; Mismatches 412; Indels 535; Gaps 55;

QY 442 TAAHNTMPTTQAPLISSEEFYREONDKGR-----SKFPYSGSSYRSSTYDSOGIAQ 495

Db 669 tslerqpkswapshepfmlmgtdsgdekdstkegptgqg--aedastes--ghye 725

QY 496 HIALTLSPSAHTLDLHDH-----PPPEAEARSAMT----- 530

Db 726 fdreihlvsskmeplkfpdpeftgmstsvspiprekeeseaeedllderkpfdvsdgv 785

QY 531 -----VAMPDLMIAHA-----RSRSPYRRYRSRSRSPREF----- 562

Db 786 ngyfervnsvverpnrllelylneetdlnaileerekeghsgdteynslpfpypren 845

QY 563 -----RGQSPTRK--NVPREKEERYFNRYREV-----PPYDIKAYVGRSV 602

Db 846 ddgyvgggkvpqitddhnpgenkd-----dykellsmntleentykppptlaastlmasr 900

QY 603 DFRDPEEK-----EKYREMRKYREWKYTKGAVGAQPPPSANREDPSPERL 651

Db 901 krtdprkrltlrsslleerygalerri-----sqdqpsgdq----- 939

QY 652 LPLNIRNSPTFRGRGRREY-----AAGQSHRNRLGNYPEKL-----SFRD--SHNA 696

Db 940 ----akypstaaleerfnlekglsaeqkqlsemaeaylpkseripstadlesrfs 995

QY 697 KDNPKSKEKESENVPGD-----GKGNKHKKRRKRNE--EKGESESPFLNPLETSR 747

Db 996 ltkgmssessskrpdlkdedrpsgsskngksektskhkeepesntketgetea 1055

QY 748 KCRGSGIDETKTD--TLFVLPS-----RDA 772

Db 1056 sdsndskigekeeteprikkipstaeldrfnalerkmsvqkspsknkpepdeesks 1115

QY 773 TPVRDEPMDAESTIFKS-----VSDDKDKREK-----KPVKSDKT 808

Db 1116 lkepeeeesekaneeksgrtqtpiakkdskdsqdksetkengspctngdeapdkpsks 1175

QY 809 KR--KSDGSATAKKDNVLRKSGKPGQKVDGDR-----KSPSEPPLKRAKKEA- 855

Db 1176 eemleketnsnpkeds--hesaatlnkvegnrelsekgdhkilekseepngksgketae 1234

QY 856 TKIDSVPKPS--SSQKDE--KVGTG--PRKAHSKSAKTRQSOQR----- 895

Db 1235 tknanvxdskkgdsgkneaaaktsvsgtesdlkpsskenstskdaeqektpkpsptee 1294

QY 896 -----TRRSKRTVPKTNSSQK--SQPVTRRRRR-----SLRKINLILIA 930

Db 1295 lekrfnalekqmtlnletkpeqtkpalksgtsaeavtkqksmfddkilevnaale 1354

QY 931 REKN-----ERERKRKKSIV-----DKDFE 948

Db 1355 kegsrveevrnaekkrknveeapknkegdsqpeesghkqngrrasepstedelekrye 1414

QY 949 SSSMKISK-----VBGTEIVKPSPKRK-----MEGD----- 974

Db 1415 tlkrmsknqfsetvdealerlqgevalseaveekpppstledlesrfealhghkknves 1474

QY 975 -----VEKLE-----RTPPKDK 986

Db 1475 kmelckhvdaleahlpstpppppkerpvlsearvlhqqalileelqskmrgsqspgeen 1534

QY 987 I-ASSTPPAKKIK----- 998

Db 1535 lkpselnprqrrklllqrpupmgdetsaeapantayraanhqvgqrmvrrfdlpsrad 1594

QY 999 -----LNRETGKI-----GNAENASTTKEPSEKLESTSSKIKQKVGAKARRVAG 1045

QY 879 KAHKSAKDRTRQSGPRTKRKRTVPTKSSQKSPVTRRRRLAKINYLIAKENEK 938
D 1430 saggaastprtrdrkrktarsr---nanseegsarknr-----gsIsakalkkrtin 1479
QY 939 RKKGV-PKDFSSSMKISKVGTETLVKSPKRMKGDEKLEPTREPKRIKXSS----- 990
D 1480 rgrlrvpeadgeddcl-----drtppepppsmdsnkrrsrtltqtkkylidvmlftsd 1534
QY 991 -----TPPAKIKILNRETGKKIGNMENASTTKEPSEKLESTSSKIKQEKVKGRKRRV 1043
D 1535 densllvaspykkdkd-----kpsanasnagsd-vekte-----pqsgaegdaagev 1581
QY 1044 AGSGSSSTLVDTY---STSSGSPVRKSEKTDTRKTVTKTMEENNDNTAEDYIT 1100
D 1582 -geeksnrlpidesqleassstlsavaekergjstdaanaasrkpruyyIntgdedsmv 1640
QY 1101 MIOVQSGWMDKDFESEBEDVKTTOPIQSVGKPSIIKKNVTKRSAKRY--TEKESQPE 1159
D 1641 qI-Vlammgkrtellldkpkkeapekpydeek-seldeaatlckpegdekltgees---- 1694
QY 1160 KIQKLPKEASHELMQHELRSKSGASSEKGRAKDREHSGSEKDPDKRK-----SGAQP 1214
D 1695 -----kkdlitseeetllessamevdskeesepd-dskksdeahkdkdkmevdeavgksd 1747
QY 1215 KESTVDRLSF---QGHFKTILSQSKETRTSEKHESVRGSSNKDFTPGRDKKVDYDSRDY 1270
D 1748 keskpgeqsetvkteenskaledkssvtladh-----akepetvlekmnevdekand 1801
QY 1271 SSSRRBERGELARRKDSPP-----RGKESL---SGOKSKIRERDLPKKAESKKSNS 1322
D 1802 qsaavsksegsektcdsdnppeatleknkesllegekevkegeevykend---te 1858
QY 1323 PPRDKP 1329
D 1859 admenkp 1865
RESULT 12
ABG12812
ID ABG12812 standard; Protein; 3660 AA.
AC ABG12812;
DT 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #12803.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001MO-US08631.
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB; AAS76999.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX
PS Claim 20; SEQ ID No 43171; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (II) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3660 AA;
Query Match 4.6%; Score 340; DB 22; Length 3660;
Best Local Similarity 20.1%; Pred. No. 2, 9e-11;
Matches 283; Conservative 153; Mismatches 463; Indels 512; Gaps 60;
QY 111 LIANKFLRQAVNNKNETGYTKRLKQLPPF-----LFLVPPRPLISQRLQ 157
D 59 llmsnyrgq-----kkgqakklekaeqgkrkvtkyknlrqfllmpvsaaisrlk 111
QY 158 PRSRPLRQODPVY-----FRYVSPCTSDPKTKAGSCSD-----SGTSLRLPAP 202
D 112 tprflededydpkrlarlestpnstfs-apsccsgksksaaqhsqmsdsstss 170
QY 203 SSSLTSNOS-----LAPVSGNPSSAPAPVDPITATVTSIVHSEKSDGPFRRSDKLL 257
D 171 svdstsdsqaseelqvipeersdtevp-h-plp-----lsgpenendrrrrys---- 220
QY 258 PAAALTSEHSGKASSIAITALMEKGVGTSPMNSIFVGSLHGLIPTTGPVRIAR 317
D 221 -----versfgrt-----
QY 318 PGGRPRGWHNSNKGIVLSPPOQIRRGERSCYRSINGRHHSERSQRTOSPLPATPCFV 377
D 231 -----lklstlqsapqg-----qlsssp----- 249
QY 378 PVPPPLYPVPP-----HTLPLPGVPP--POFSPQFSSQPTAGVS--VPPG 423
D 250 --ppplltppplqpassisdht--pwlmpcltplasflpaspatamgkkrksllrept 304
QY 424 F-----PPAPANISTAGFS-PGV--PTAHSNTPMTTOAPLISREEPYREONDKRES 472
D 305 frwtslkhsrsepqyfsesakyakeglirxpifdfrppltp-----edvgfas 353
QY 473 KFPYSSGSSYRSSTYDSSQGLAQHIIHALTILSPSAHRLDLHDHPRPEEAARSAMIVH 532
D 354 gfsasglaaaraalfsplhsgrlfomh--krs-----llrprtfpsaahsllfesvt 404
QY 533 MPDLMDIAHARSRSPRYRRRSRSRSPPEFGOSPTRKNVPRREEKREYNRYREVPVY 592
D 405 lpsnrtsaagcsgsvmrkkrkvfslrseprps--hsmrtsgslsselspltp 462
QY 593 DIRAYYGRSVD-----FRDPEKERYREMERKYREMYKYGYVGAQPRP--- 639
D 463 svsselslsvslatsalnpltfps-----hsltgsgeaeknqrprk 507
QY 640 -SANREDF-----SPRLLPLMIRNSPFTGRGRREDVYAGOSHRNRLGNGYREKLS-TRDSH 694


```

Db      508 tsapaeffssssplrlpfwftpgsqtergrnkda-----peelckdrdad 553
QY      695 NAKNPPSKKEESENVPDGGKGNKHKHRRRNEKEGESESLNP-----ELL 743
Db      554 ksvckdksrerdrer-----ekenkresrkekrkgselqssalypgrvskvkgadv 609
QY      744 ETSKCKGSSGIDETKTDTLFVLPSRDDATPVREPMASITKSVSDKKRKKDKV 803
Db      610 atssasakalgrkss-----shdsqf-----dltsvlt----- 638
QY      804 KSDKTKRRSDSATAKKDNVL-KPSKGPQEKVDGDRKSPSEPLKKAKE---EATKID 859
Db      639 -----gdtavatklikkgrnlektnd--lgr-tapsleklekclclstpsss 685
QY      860 SVKPSSSS-----OKDEKVTGTPRKANSKAKDTRROSOPTRRSKRTVPTS 907
Db      686 tvkhstssisglmaqdklpmtdkrvasllkkakqclcklsksklqtdg----pkaq 740
QY      908 SOKSOPVAT--RRPRSLRKINYLAREKNERKRRKSVDKOFESSSMKISKVEGTETVKP 965
Db      741 gqesdssetsvrgr-----lkh-vcirraavalytkravfpd----- 776
QY      966 SPKRKMEGVKLEBRTP--EKDKIASSTTPAKKIKILNRETCKKIGNAENASTTEPSEKL 1023
Db      777 -----dmpltsalpwereekllss-----mgndksslagsedaaplpi 820
QY      1024 ES-TSSKTKOKK--VKGAKAKRVAGSEB-----SSSTLVDTYSTSSGSGPVKRSSEKTD 1075
Db      821 kvprtnkqgepvvkkgrsrrcgcqpcqyvedcgvctnldkpkfigrnlkkqckc-- 878
QY      1076 TKRTVTKTMEYNNDNTPAPADVILIMIQVPSKMDKODFESEEDPVKTQPIQSVGKPS 1135
Db      879 -----mrkcnlqwmprskaylqkqakavkkkkekstkckd-----skess 920
QY      1136 IIKNVTKPSATARYTEKESDOPKLOKLPKASHLMOHELRRSSGKSASSEKGRAKDRE 1195
Db      921 vvkhnv-----dsqgkp----- 932
QY      1196 HSGSEKNDPDKRKSGAODKSTYDRLSEOGHFKTLSOSSKETRSEKHESVKSSNKDF 1255
Db      933 -tparedpakpkassepprkpvceeksegnvasapgesqatparcksskqvsqpal 991
QY      1256 -----TPGRDKKVDYSDRYSSSKRRDERGELARRKDSPPRKESLSGQKSKLREERD 1308
Db      992 vlpqppttpprk-----evpkltpsepkkkp-----p 1021
QY      1309 LPKKAESKKSNSPPRDKKPHDKAPYETKRPCBETKRPVKD-----NS 1352
Db      1022 ppsqsgpqskqkvaprpstlvkqk-pkekekp-----ppvnkgenagtlmlstlsngns 1076
QY      1353 GKER-----EKHAAPARKNGESSG 1371
Db      1077 skqkfpadgyvhrivdfkedceaanvwemg 1107

RESULT 13
ID      AAM78825 standard; Protein; 1026 AA.
XX      AAM78825;
XX      06-NOV-2001 (first entry)
XX      Human protein SEQ ID NO 1487.
XX      Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX      vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
XX      tissue growth factor; immunomodulatory; cancer; leukemia;
XX      nervous system disorder; arthritis; inflammation.
XX      Homo sapiens.
XX

```

```

PN      W0200157190-A2.
XX
PD      09-AUG-2001.
XX
PE      05-FEB-2001; 2001WO-US04098.
XX
PR      03-FEB-2000; 2000US-0469614.
PR      27-APR-2000; 2000US-0560875.
PR      20-JUN-2000; 2000US-0598075.
PR      19-JUL-2000; 2000US-0620325.
PR      01-SEP-2000; 2000US-0654936.
PR      15-SEP-2000; 2000US-0663561.
PR      20-OCT-2000; 2000US-0693325.
PR      30-NOV-2000; 2000US-0728422.
XX
PA      (HYSE-) HYSEQ INC.
XX
PI      Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI      Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI      Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX      WPI: 2001-476283/51.
XX      N-PSDB: AAK51958.
XX
PT      Nucleic acids encoding polypeptides with cytokine-like activities,
PT      useful in diagnosis and gene therapy -
XX
PS      Claim 20; Page 3768-3770; 6221pp; English.
XX
CC      The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC      encoded polypeptides (AAM78323-AAH80302) that exhibit activity elating to
CC      cytokine, cell proliferation or cell differentiation or which may induce
CC      production of other cytokines in other cell populations. The
CC      polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC      peptide therapy. The polypeptides have various cytokine-like activities,
CC      e.g. stem cell growth factor activity, haematopoiesis regulating
CC      activity, tissue growth factor activity, immunomodulatory activity and
CC      activin/inhibin activity and may be useful in the diagnosis and/or
CC      treatment of cancer, leukemia, nervous system disorders, arthritis and
CC      inflammation.
CC      Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC      (AAM80020) are omitted as the relevant pages from the sequence listing
CC      were missing at the time of publication.
XX
SQ      Sequence 1026 AA:

Query Match          4.5%; Score 335.5; DB 22; Length 1026;
Best Local Similarity 21.8%; Pred. No. 1e-11;
Matches 189; Conservative 142; Mismatches 341; Indels 193; Gaps 38;

QY      543 RSRSPRYRYSRGRS-PPEFRGOSPTKRNVPREKEKEYNRNKREVPVPPIDIKAYYRS 601
Db      312 rsageeleryrrqlqartlealeakstkdslerqsele--drbqa-----di-asyqea 363
QY      602 VDFRDPFEKERYREWE-----RKYREW-----YEKYVKG--YAVGQPRRSA 641
Db      364 igqld--aelnrtkwemaagltreayqdlinvkmaldieaayfklllegeccrfigrpp-- 419
QY      642 NREDFSPERLLPLNIRNSPFTGRRED--YAAGOSHRRNRNLGNYPEKLISTRDSHNADN 699
Db      420 ----fslpegiplkpsvsthklyvkeekelkvekeketivveeqteetqyleveetee 475
QY      700 PKSKKESEBNVPGDGKGNKHKHRRRNEKEGESESLNPPELLFTSRCKCGSSSIDETK 759
Db      476 keakeeg-----keeggeee-----eggeetk 502
QY      760 TDTL--FVLPSRDOTAPVDE---PMDAESTFKFSVSKDKREKDKPVKVSUKTRKSDG 814
Db      503 spvveaaspeakeaspvkeekspakeas-----pekeekaspeevkspeakspaspa 554
QY      815 SATAK-----KDNVLKPSKGPQEKVDGDRKSPRSEPLKRAKEBATKIDSVKPSSSQK 869

```



```

Db      891  decasqnevtv-epqtladsvpdiapakemdtledexvtnveenkseesekpenpee 949
Qy      1168  ASHELMQ-----HELRSSKGSASSEKGRANDREHSGSEKNDNPKRKGAGQ----- 1212
Db      950  cvdepaqvqdvadppqdaakpatvpvpinseqdfeadivthfeentdeftrliqlinql 1009
Qy      1213  -PDKESTVDRLSEOG-----HFETLSQSSKETRTSEKHESVSGSSNKDFTPGRDKKDY 1265
Db      1010  ledrknlnrlsedgtqeeavdlralrlrslskrrrsmqeqhr--veretlppqprpsps 1067
Qy      1266  DSRDYSSSK--RRDERGELARRKDSPPRGKESLSGOKSKLREERD 1308
Db      1068  sstagspakrllrqde-----pkssplpsdaslnskenealekqe 1106

```

Search completed: September 12, 2002, 18:28:42
 Job time: 11242 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2002, 16:03:56 ; Search time 25.89 seconds
(without alignments)
1324.586 Million cell updates/sec

Title: US-09-811-045A-1
Perfect score: 7374
Sequence: 1 MWEVKDPMKGMGLNTGKY.....WRSWLGRRWRPSSRNP 1404

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata1/1/aa/5A.COMB.pep:*
2: /cgn2_6/ptodata1/1/aa/5B.COMB.pep:*
3: /cgn2_6/ptodata1/1/aa/6A.COMB.pep:*
4: /cgn2_6/ptodata1/1/aa/6B.COMB.pep:*
5: /cgn2_6/ptodata1/1/aa/PCPUS.COMB.pep:*
6: /cgn2_6/ptodata1/1/aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	340	4.6	3969	4 US-08-061-376-5	Sequence 5, Appl1
2	276.5	3.7	1187	1 US-08-320-559-28	Sequence 28, Appl1
3	276.5	3.7	1187	3 US-08-545-860D-28	Sequence 28, Appl1
4	276.5	3.7	1187	5 PCT-US94-04496-28	Sequence 28, Appl1
5	276.5	3.7	1210	1 US-08-320-559-26	Sequence 26, Appl1
6	276.5	3.7	1210	3 US-08-545-860D-26	Sequence 26, Appl1
7	276.5	3.7	1210	5 PCT-US94-04496-26	Sequence 26, Appl1
8	274	3.7	2842	1 US-07-741-940-7	Sequence 7, Appl1
9	274	3.7	2842	1 US-08-289-548A-7	Sequence 7, Appl1
10	274	3.7	2842	1 US-08-452-654-7	Sequence 7, Appl1
11	274	3.7	2843	1 US-08-452-655B-2	Sequence 2, Appl1
12	274	3.7	2843	1 US-08-452-655B-7	Sequence 2, Appl1
13	274	3.7	2843	3 US-08-450-582-2	Sequence 2, Appl1
14	274	3.7	2843	3 US-08-450-582-7	Sequence 7, Appl1
15	274	3.7	2973	2 US-08-821-355A-7	Sequence 7, Appl1
16	274	3.7	2973	2 US-09-003-687A-7	Sequence 7, Appl1
17	274	3.7	2973	4 US-09-136-605-7	Sequence 7, Appl1
18	273.5	3.7	1588	5 PCT-US93-07261-11	Sequence 11, Appl1
19	273.5	3.7	1663	5 PCT-US93-07261-16	Sequence 16, Appl1
20	273.5	3.7	1780	1 US-08-769-309A-5	Sequence 5, Appl1
21	273.5	3.7	1780	1 US-08-769-309A-5	Sequence 5, Appl1
22	272.5	3.7	1162	2 US-08-728-323A-2	Sequence 2, Appl1
23	270	3.7	2843	1 US-07-741-940-2	Sequence 2, Appl1
24	270	3.7	2843	1 US-08-289-548A-2	Sequence 2, Appl1
25	270	3.7	2843	1 US-08-452-654-2	Sequence 2, Appl1
26	270	3.7	2843	1 US-08-370-235A-2	Sequence 2, Appl1
27	264	3.6	688	3 US-09-141-047-8	Sequence 8, Appl1

28	262	3.6	1898	1 US-08-056-200-94	Sequence 94, Appl1
29	262	3.6	1898	2 US-08-800-644-94	Sequence 94, Appl1
30	261.5	3.5	1581	4 US-09-110-517-2	Sequence 2, Appl1
31	259	3.5	434	1 US-08-097-830E-3	Sequence 3, Appl1
32	259	3.5	434	2 US-08-456-112B-3	Sequence 3, Appl1
33	250.5	3.4	1706	2 US-08-459-568-2	Sequence 2, Appl1
34	250.5	3.4	1706	2 US-08-399-411-2	Sequence 2, Appl1
35	250.5	3.4	1706	3 US-08-516-859A-2	Sequence 2, Appl1
36	241	3.3	1400	1 US-08-080-255-7	Sequence 7, Appl1
37	241	3.3	1400	1 US-08-465-713-7	Sequence 7, Appl1
38	241	3.3	1400	5 PCT-US93-05857-7	Sequence 7, Appl1
39	235	3.2	683	6 5210183-3	Sequence 7, Appl1
40	226	3.1	885	2 US-08-533-306A-4	Sequence 4, Appl1
41	226	3.1	885	2 US-08-742-923A-4	Sequence 4, Appl1
42	225	3.1	1719	2 US-08-459-568-4	Sequence 4, Appl1
43	225	3.1	1719	2 US-08-399-411-4	Sequence 4, Appl1
44	225	3.1	1719	3 US-08-516-859A-4	Sequence 4, Appl1
45	224	3.0	571	4 US-08-961-083-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-08-061-376-5
; Sequence 5, Application US/08061376
; Patent No. 6175000
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; APPLICANT: Djabali, Malek
; APPLICANT: Selleri, Lucia
; APPLICANT: Parry, Pauline
; TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11023
; NUMBER OF INVENTION: TRANSLATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/061,376
; FILING DATE: 13-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9387
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)546-4737
; TELEFAX: (619)546-9392
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3969 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-061-376-5

Query Match 4.6%: Score 340; DB 4; Length 3969;
Best Local Similarity 20.1%: Pred. No. 9.5e-14;
Matches 283; Conservative 153; Mismatches 463; Indels 512; Gaps 60;
QY 111 LIANKFLQAAVNNKFNKGTGTRLRKQLPPF-----LFLVPPRPPLSRNLQ 157

```

Db 370 LLSNSYRGQ-----KKGAQKKIEKAAOLOGRKVKYQYKNIROFIIMPVSAISSRIIK 422
QY 158 PRSRPIRLQODPV-----FRYTVSPICSDTKTGSCSD-----SGLSRLPAP 202
Db 423 TPRRFIEDEYDPPIKARLESTPNRHS-APSCGSEKSSAASQHSQMSQSSRSP 441
QY 203 SISLTSNOS-----LAPVSGNPSPAPVPDITATVSIYHSEKSDPFRSDNKL 257
Db 482 SVOTSDQASAEIOVLPERSDTPVHP-PLP-----ISQSPENESNDRSRHS- 531
QY 258 PAAALISEHKGASIAITRLMEKGVPGTSPMNSITVGOSLIHQILPTGPRINAR 317
Db 532 -----VSESRFGSRT----- 541
QY 318 PGGGRPGMEHNSKLYTVSPPOOIRGERSCYRINNGRHRSRQRTOSPLPATPCFV 377
Db 542 -----TKLSTLOSAPOQ-----QTSSP----- 560
QY 378 PVPPPLPYPP-----HTLPLPGVP--POFSPQFPSSQPTAGS---VPPPG 423
Db 561 --PPPLTTPPPLOPAPASSIDHT--PWLMPPTIPLASPLPASTAPMQRKRSILREPT 615
QY 424 F-----PPRANISTACFS-PCV--PTAHNTMPTTAPPLISREPTREONDGREG 472
Db 616 FRMTSLKHSRSEPOYFSSAKYAKEGLRKPIFDNFRPPLTP-----EDVGFAS 664
QY 473 KFPYSGSSSYRSDSSOGLAOIHAIHLTLSPAHTLDLHDPHPPEAEARSAIVH 532
Db 665 GFASGTAASARLFLPSLHSGTRFDMH--KSP-----LLRAPRTPSAHRSIESVT 715
QY 533 MPDLMDIAHARSPPRYRHSRSRSPERFGOSPTKRNVPREKEREYFNRYREVPY 592
Db 716 LPSNRTSAGTSSGVSNGRKRKRYFSPIRSEPRSP--HSMRTSRGLSSELSPLPPS 773
QY 593 DIRAYYGRSD-----FDPPEKERYREMERKYREMYEKYGVAGQPRP--- 639
Db 774 SVSSSISIVSPATSAALNPTFFPS-----HSLTOSGEAEKRNQRRKO 818
QY 640 -SANREDF--SPERLLPLNIRNSPPTGRREDYAGQSHRNRLNGVPEKLS-TPDSH 694
Db 819 TSAAPAEFSSSPPLPFWPTPSSOTERGKNNKA-----PEELSKDRAD 864
QY 695 NAKONPKSKESSENVPGDGKNNKHKRRNRNEKESESEFLN-----ELL 743
Db 865 KVEKDKRSREDRER--EKENKRESRKRKRGSEIGSSALYPGVSKEKVEDEV 920
QY 744 ETSRKGCGSSGIDETKTDTLFLVLRSDDATPVDEPMDAESTIFKVSXDKREKDKPV 803
Db 921 ATSSSAKKAIGRKSS-----SHDSGT-----DITSVTL----- 949
QY 804 KSDTKRKSDSATKAKDNVL--KPSKPOEYVDGREKSPRSEPPLKAKE-----EATKID 859
Db 950 -----GDTTAVTKTILIKKGRONLEKTILD--LGP-TAPSLKEKTKLCLSPSS 996
QY 860 SVKPSSS-----QKDEKYTGPRKAHSAKDTROSQPRTRRSKRYPKTS 907
Db 997 TVKHSTSSISGMLAQADKLPMWTDKRVASLTKAKAQLCKEKSKSLKQTDQ-----PKAQ 1051
QY 908 SQASQVPR--RRPSLRKINYLIAREKNERERKKSVDKDFSSSMKISKVEGTIVK 965
Db 1052 GQSDSSETSVGRPR--IKH-VCRRAAVALGRKRAVPDP----- 1087
QY 966 SPRKMEGVEXELTRP--EKDKIASSTTPAKKIKLNRGKKIGAENAASTKEPSEKL 1023
Db 1088 -----DPTLSALPWEERKILSS-----MGNDKSSIAGEDEPLAPPIKPI 1131
QY 1024 ES-TSSKIKOEK--VKGAKRKVAGSEG-----SSSTLVDTYSTSGSPVSKSEKTD 1075
Db 1132 KPTVTRKKAPOEPVKKGRSGRGOCPCGCVPCNCIDKPKFGGNIKKQCK-- 1189
QY 1076 TKRTVKTMEEVNNDTAPAEVYIMIQVPOSKWMDDESEEDVKTQPIQSVKPS 1135

```

```

Db 1190 -----MRKCONLQMPESKAYLOKQANAVKKREKKSRTSEKDD-----SKESS 1231
QY 1136 IIRKNTTKSATAKYTEKESEOEKLOKLPRKASHELMQHELSSKGSASEKRAKORE 1195
Db 1232 VKNVNV-----DSSQKP----- 1243
QY 1196 HSGSEKDNDRKRSQAOPKESTVDRLSEOGHFKTLSQSKETRTSEKHEVSGSSNKP 1255
Db 1244 -TPSAREDPAPKSSSEPPPKRYEKESEBGNVASGPESKQATTPASRKSQVQPAL 1302
QY 1256 -----TGDRKKVYVDSNDYSSSKRDERGELARKDSPPRGKESLSGQSKLREED 1308
Db 1303 VDPQPPPTGPPRK-----EVPTTSEPPRKQP-----P 1332
QY 1309 LPKGAESKSSNSPPRDKPHDKAPYETKRPCEETKPVDK-----NS 1352
Db 1333 PPSGEPQSKQKAVARPSPVPKQ--PKERKP-----PPVKNQENAGTILSTLSNGNS 1387
QY 1353 GKER-----EKHAERANGKESSG 1371
Db 1388 SKQKIPADGVHRIYDFKEDCEAENWEMMG 1418

RESULT 2
US-08-320-559-28
; Sequence 28, Application US/08320559
; Patent No. 5633135
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
; TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135rlis
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/320, 559
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/062,443
; FILING DATE: 14 MAY 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/971,094
; FILING DATE: 30-OCT-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,830
; FILING DATE: 27-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/805,093
; FILING DATE: 11-DEC-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-0855
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1187 amino acids

```

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-320-559-28

Query Match 3.7%; Score 276.5; DB 1; Length 1187;
Best Local Similarity 18.3%; Pred. No. 3.2e-10;
Matches 227; Conservative 160; Mismatches 428; Indels 423; Gaps 55;

QY 231 DITATVSVISVHSEKSDGPFSDNKL-LPAALITSEHSGASSIAITALEMEKGVPGTSP 289
DB 43 EVKEFLSTKSHTRLDA-----SENRLGPKYPLIPD-KGSS-----IPSSSF 84
QY 290 WNSIVVGSGLHGOLI-PTTGPRVIRIN-AARPGGGRGWE-----HSNKLGLVLSPPQ- 339
DB 85 HTSV-----HQSHHTPASGPLSVGNISHNPKMAQPTPEMPSLIHAKSCG-----PPDS 133
QY 340 ----QIRGERGCYSINGRHHSERQSOTGSLPATPCFVPVPPPLYPHPHTLPLP 395
DB 134 QHLTDRIGQEGFGSS-----HHKGDRRADG-----DHCASTY 167
QY 396 PGVPPPOESP-QFPSSQPTAGYVPPGPPAPANISTACSPGVPTAHSN-TMPTT 451
DB 168 DSAPFERELSLTSLSPVPP-----LSP-----IHSNQTLPT 201
QY 452 QAPLSREFYEQNDKGRESKFPYSGSSYSRST--TDSSGGLAOHIH-----A 499
DB 202 QG-----SSK--VHGSSNNKGVCYCPAKSPKDLAVKVDKETPODSLVA 242
QY 500 LTLSPSAHTLIDLHDPHPPEAEARSAMIYHMPDLMDIAHARSPPRYHRSRSP 559
DB 243 PAQPS-----QTFPPSLPSKSYAMQKP-----TAYVRMD-----GQDAAP 281
QY 560 PEFQSGPTKRNVPREKEREYFNRYREVP-----PPYDIKAVYGRSVDERDPE 609
DB 282 SE-----SEELKLPEDYEQTEKTDLKVPAAKLTLLKMPQSGVEQYTSNEHCVEELL 337
QY 610 KERYEMEKRYEMYEKYYKGYAVGAQPPPSANREDPSERLPLNIRN-----SPTRR 665
DB 338 KEMTSHW-----PPPLTAIHTPTSTAPSKFPPTPKDSQHVSVTQMO 379
QY 666 REDYAGOSHRNRNLGN--TYPEKLSTRDSHNAKNPKSEKESENVPDGKGNHKKHR 723
DB 380 KQYDTSKTHSNSQGTSMLEDDIQLSD-----SEDSDSQTP----- 418
QY 724 KRNEKEGESESLNPELL-ETSRCKGSSGIDETKTDTLFLVLSRDDATVDERPMDA 782
DB 419 ----EKPPSSSAPPSAPQSLPEPVASAHSSSAESESSTDS-----DSSSDS 460
QY 733 ESTFKSVSDKDK-REKDKPKYKSDKTRKKSQDSATAKKDNVLSKGPQEKVDDREKS 841
DB 461 ESESSSSDSEENPLETPAPEPEPTTNWQJDNMLTKSQPAAPPEGGRS-----TEP 514
QY 842 PRSEPLKAKKEATKIDSVKPPSSSQKDEKVTGTPRKASHSKANDTRQSPRTRRSKR 901
DB 515 PRHPEKSGSSSAT-----SQEHSEKDPKPPKSSKAPAPAPAPAP-----GKR 560
QY 902 TVPKTSSQSKQVYR-----TRPRSLKINYLIAREKNEREKKKSVKDKDESSSKISKIV 957
DB 561 SQCKSPAOQEPQROTGVGKQPKPKV-----ASAPAGSRSTSLQGREREGLLPYGSR 612
QY 958 ECTEIVKPSPKRMGDEVEKLETRPEKDKIASSTTPAKKIKINRELGKIGNAENASTJK 1017
DB 613 DOTS-KDKPKYKTG-----RPRAAASN 634
QY 1018 EPSEKLESTSTKIQOEKYGKAKRRVAGSEGSSTLVDT-----STSGTGSFVRS 1070
DB 635 EKPAVPAPSESEKKKHSLSLAPSKALSGEPAKDNVEDTPEHFAVLVLTESOGPRHSGS 694
QY 1071 EKTDTKRVIVITMEYINNDNTAPADVDVIMIQVPOSKWKDKDFSESEEDVKTPTDIOGV 1130
DB 695 GSRTSGCROAVVVOEDSRKDR-----LPPLP-----NDTKLISPLRDT 732

QY 1131 GKPSSTINVTTPKSPATAKYTEKESQPEKLOKLPRKASHELMOHELSSKGSASEKGR 1190
DB 733 PEPQSLMWKITL-----DLISRIOP-----PGKG---SRQK 762
QY 1191 AKDRHSGSEKDNPDKRRSGAOPDKESTYVDRLSEQGHFKTLSQSKETRTPSKHBSVGS 1250
DB 763 AEDKQPPAGKHHSEKRS-----SDSSKILAKRRKGAEKDC 799
QY 1231 SNKDETPGRDKKVDYDSRDYSSSKRRDERGELAR-----RKDSPPRKEESLSGQ---- 1300
DB 800 DNKKIR--LEKEIKSQSSSSSHKESKTKPSRPSQSSKEMLPAPPVSSSQKPAKP 857
QY 1301 --SKUREDL-----PKGAEKSN---SSPPDK-----PHDK-APYETKRP- 1341
DB 858 ALKRSRRADTCGODPPKASSTKSNHDKDSSIPKQRYVEGKGRSSSEHKSSGDTANPF 917
QY 1342 -----CEETKPVDNKSGKEREKHAHAENKRESS 1370
DB 918 PVPSPNGNSKRGKQVYKFDKQADLHMRKAKMKOKA 955

RESULT 3
US-08-545-860D-28
Sequence 28, Application US/08545860D
Patent No. 6040140
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: CanaanI, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1 Region
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
ADDRESSEE: No. 6040140ris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545, 860D
FILING DATE: 07-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE: 22-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10930
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/327,392
FILING DATE: 19-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/320,559
FILING DATE: 11-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,443
FILING DATE: 14-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,094
FILING DATE: 30-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/888,839
FILING DATE: 27-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,093

```

QY 958 EGTETLVKPSPRKRMKGVDKELTEPTEPKDKIASTTPRAKIKILNREIGKIKGNAENASTTK 1017
Db 613 DOTS--KKPKPKYTG-----RPRMAASN 634
QY 1018 ESESELESTSSKIKQEKVKGAKRKAVSGSGSSSTLVDT-----STSTGSGSPVRKS 1070
Db 635 EKPRAVPSPSEKKKHKSSCLPADPSKALSGEPRAKDVEDTPEHFALVPLTESGGPPHSGS 694
QY 1071 EKKTKTKTVITMEEYNDNTAPARADVILIMIQVPSKWKDDFESEEDCYKTPQISY 1130
Db 695 GSRTSGCGRAVVVQEDSRKDR-----LPLPL-----RDTKLKSLPLRDT 732
QY 1131 GKPSSTIKKVVTKPSATAKYTESQPEPKLQKLPREASHELMHQHELRRSSGASASEKGR 1190
Db 733 PPSQSLMVTTL-----DLNRIPOP-----PEKG--SKRK 762
QY 1191 ADRHESSGSEKNPDRKSGAOPDKESTVDRLSEOGHFKLTSQSKETRTSEKHESVRGS 1250
Db 763 AEDKQPPACKKHSSERS-----SSSSSKLAKKRKGELEROC 799
QY 1251 SNKDFPPGRDKKDYDSRDYSSSKRRDERGELAR-----RKDSPPRGESLSGOK--- 1300
Db 800 DNKKIR--LEKEIKSGSSSSSHKSSKTKPRSPSSQSKKEMLPPPPVSSSSOKRAPK 857
QY 1301 --SKLEERDL---PKGAESKKS--SSPRDK-----PHDK-AVYETKRP- 1341
Db 858 ALKRSRREADTCGDDPPKSSSTKSNHKSSIKQKRVESGKSGRSSSEHGSSGDTAMP 917
QY 1342 -----CEFTKPYDKNSGKERENHAARNGKESS 1370
Db 918 PVPSLPNGNSKPKGPVKYKPDQOADLHMEAKKKKKOKA 955

RESULT 4
PCT-US94-04496-28
Sequence 28, Application PC/TUS9404496
GENERAL INFORMATION:
APPLICANT: Cioce, Carlo
APPLICANT: Canaan, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESSES:
ADDRESS: Woodcock, Washburn, Kurtz, Macklewitz &
ADDRESS: Norris
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca Esq, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-1242
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3439
TELEFAX: (215) 568-3100
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1187 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

```


PCT-US94-04496-28

Query Match 3.7%; Score 276.5; DB 5; Length 1187;
 Best Local Similarity 18.3%; Pred. No. 3.2e-10;
 Matches 227; Conservative 160; Mismatches 428; Indels 423; Gaps 55;

QY 231 DITAVISIVHSEKSDGPRDSDNKL-LPAALITSEHSKASSIATLAMEKGVGTSP 289
 DB 43 EKEFLSTKSHTRUDA-----SENRLGKPKYPLIPD--KGSS-----IPSSSF 84
 QY 290 WNSIFGOSLHGLI--PTGVPVIRIN--AARPGGPRGME-----HSNKLGIYVSPPO- 339
 DB 85 HNSV-----HHQSHIFASGPLSYGNISHPKMAQPTETPMSLMAHSCG-----PPDS 133
 QY 340 ---QIRGERSCYRINRGHHSERSQRTQSDSLPATPCFVVPVPPPLYPPPTLPLP 395
 DB 134 OHLTDRLOEGFGSS-----HHKGGRRADG-----DHCAST 167
 QY 396 PGVPPQFSP--QFSSOPPTAGYSVPPGPPAPANISTACSPGVTANSN--TMPTT 451
 DB 168 DSAPEKELSLPLSPVPV-----LSP-----IHSNQTLPRT 201
 QY 452 QAPLISREFFREONDKGRESKFPYSGSSYSRSSY--TDSOGIAQH--A 499
 DB 202 QG-----SSK--VHGSSNNSKGYCPAKSPKDLAKVHDKETPODSLVA 242
 QY 500 LILSPSAHTLIDLHPPEBEAKRSAMIYHMDLMDIAHARSSPPYRRYRSRSP 559
 DB 243 PAQPPS-----QTPPPSLPSKYVAMQKP--TAYVRPM-----GQDAP 281
 QY 560 PFRGQSPTRKRVPREKREYFNRYEVP-----PPYDIAYYGRSVDPRPE 609
 DB 282 SE-----SPLKLPEDYKQOTEKIDLVAKAKLTKLMPQSOSVQTYSNEYHCVEILL 337
 QY 610 KERYREMERKYREWEKYYKGYAVGAOPPSANREDESEPERLLPLNIRN-----SPTGR 665
 DB 338 KEWTHSM-----PPPLTAHTPSTAEPSKFPPTKQSHVSVTQNO 379
 QY 666 REDYVAGOSHRNRNLGN--YEPEKLTROSHNAKNPKSKSEKSEVPGDGKGNKKHHR 723
 DB 380 KOYDTSSKTHSNQOSTSMLEDDQLSD-----SEDSSEQT 418
 QY 724 KRNEKEGESESEFLNPELL-ETSRKCRGSSGIDETKDTLFLVPRDATTVPYRDEPMA 782
 DB 419 -----EKPPSSAPPAPOSILPEPVASAHSSSAESTSDS-----DSSSDS 460
 QY 783 ESTTFKSVSDDKR-REKDKPKVSKDKTKRSDGATAKDNVLPKSGQOEKVDGDRKS 841
 DB 461 ESESSSDSEENPLETTPAPEPEPTTNKQOLDNMLTKVSQPAAPPEGRS-----TER 514
 QY 842 PSEEPPLAKAKEBATKIDVKKPSSSQKDEKYTGTPRKASHKSAKTOROSOPRTRSKR 901
 DB 515 PRHHPSPKSSSSAT-----SQEHSESKDPPPKSSSKAPAPPEAPH-----GKR 560
 QY 902 TWPKTSOSKOPR-----TRRPSLTKIYVILAREKNEREKRSKXDKOFESSMKISKY 957
 DB 561 SCOKSPAQOEPPOROTGTGKOPKPKVK-----ASARAGSTSLQGEPELPLYYGSR 612
 QY 958 EGTETVKKPPKRMEDGVKLEETPEKRIASSTTPAKKIKILNRETKIGNAENASTTK 1017
 DB 613 DQTS--KOKPKYKTKG-----RPRAAAN 634
 QY 1018 EPSEKLESTSSKIKOKKVGAKARKVAGSESSSTLVDT-----STSSGSGSVKRS 1070
 DB 635 EKPAPVPPSEKHKHSSLPAPSKALSGPEPAKDNVEDRTPHEFALVPLTESGPPHSS 694
 QY 1071 EKTIDKRTVITMEYNNDNAPADYIMIQVPQSKDKDFEEDDVTKTQIQSV 1130
 DB 695 GSRTSGCRAVVQVEDSKRDR-----LPPL-----BDTKLSPRLRT 732
 QY 1131 GKPSIIRKVVTKPSATAKYTEKESQPEKLOLPEKASHELMQHLRSKSGSASEKGR 1190

DB 733 PPQSLMVKITL-----DLISRIPQ-----PGK-----SRQK 762
 QY 1191 AKDREHSGSEKONPDKRSKGAQPDKESTYDRLSEOGHFTLSOSSKERTSKHHSVGS 1250
 DB 763 AADKOPPAKKHSSKRS-----SDSSKLAKKKKGAEKRD 799
 QY 1251 SKNDFTPGDKKVDSDRQYSSKRDGERLAR-----RKDSPRGKESISGOK----- 1300
 DB 800 DNKKIR--LEKETKSQSSSSSHKSSKTKPFRSSOSSKKEMLPPIVSSSQKPAK 857
 QY 1301 --SKUNEERDL-----PKGAEKRSN--SPPRDK-----PHDK-APYETKRP- 1341
 DB 858 ALKRSREADTGGDPPKASSTKSNHKDSILPKORVVGKSRSSSEHKSGSGDTANF 917
 QY 1342 -----CEETKPYDKNSGKREKHAALANKGESS 1370
 DB 918 PVPSLPNGNSKPKGPQVKFEDKQOADLHMEAKMKOKA 955

RESULT 5
 US-08-320-559-26
 ; Sequence 26, Application US/08320559
 ; Patent No. 563135
 ; GENERAL INFORMATION:
 ; APPLICANT: Croce, Carlo
 ; APPLICANT: Cnaan, Eli
 ; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
 ; TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
 ; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
 ; NUMBER OF SEQUENCES: 44
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 563135Tis
 ; STREET: One Liberty Place - 46th Floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: PC-DOS
 ; SOFTWARE: WORDPERECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/320,559
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/062,443
 ; FILING DATE: 14 MAY 1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/971,094
 ; FILING DATE: 30-OCT-92
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/888,830
 ; FILING DATE: 27-MAY-92
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/805,093
 ; FILING DATE: 11-DEC-91
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Deluca, Mark
 ; REGISTRATION NUMBER: 33,229
 ; REFERENCE/DOCKET NUMBER: TJU-0855
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (215) 568-3100
 ; TELEFAX: (215) 568-3439
 ; INFORMATION FOR SEQ ID NO: 26:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1210 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-320-559-26

Query Match 3.7%; Score 276.5; DB 1; Length 1210;
 Best Local Similarity 18.3%; Pred. No. 3.3e-10;
 Matches 227; Conservative 160; Mismatches 428; Indels 423; Gaps 55;

```

Oy 221 DTTATVSVSHSEKSDGPRDSNKL-LPAAALTEHSHKASSIAITALEMEKGVGTSP 289
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 EVKEFLSTFKSHHRLDA-----SENRLGKPKYLPID--KGSS-----IPSSSF 107

Oy 230 WNSIVVGSGLHGOLI--PTTGPRVILN--AARPGGGRGWE-----HSKKLGIVLSPQ- 339
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 108 HTSV-----HHQSHHPPASGPLVGNISHNPKMAQPTTEPMPLIAKSCG----PPDS 156

Oy 340 ----QIRRGERSCYRSINRGHRSERSOPTSPSLPATPCFVVPVPPPLYPPTPLPLP 395
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 137 QHLTODRLGQEGFGSS-----HHKGDRAAD-----DHCAST 190

Oy 336 PCVPPQFSP--QFPSSQPTTAGYVPPPGPPAPANISTACFSPGVPTAHSN--TMPTT 451
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 DSAPREELSLISLPSVPP-----LSP-----IHSNQTLPRT 224

Oy 452 QAPLISREFFYEQNDKREKFPYSGSSYSRST--IDSQGLAQHIN-----A 499
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 225 QG-----SSK--VHGSSNNGKGYCPAKSPKDLAVKHDKETPDQSLVA 265

Oy 500 LTLSPSAHTLDLHDHPHPEAEARSAMIYHMPDLMDIAHARSRPYRKYRSRSP 559
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 266 PAQPS-----QTFPPSLPSKSYAMQKP-----TAYVRMD-----GDDQAP 304

Oy 560 PEFQSQPTKRNVPREKEREYFNRYREVY-----PPYDIKAYGSRVDEPRPE 609
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 305 SE-----SPELKLPLPDYKQOETKTDLKVPAKAKLTLLKMPQSVGEQYVNEHVCEIL 360

Oy 610 KRYREMEKRYEMWEKYYKGAVAQAQPPRSANREDFESERLLPININ----SFTGR 665
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 361 KEMTSHW-----PPLTAIHPTSTAPSKFPPTTQDSQHVSSVTQO 402

Oy 666 REDYAGOSHRRNRLGN--YPEKLSTRDSHNAKDNPKSEKESENVGDGKGNHKKHR 723
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 403 KQDYDSKTHSNQSGTSMLEDDQLSD-----SEDSDSQOTP-----441

Oy 724 KRNEKEGSESLNPELL-ETSRCKRSGSIDETKTDTLFVLPSRDDATPVREPDMA 782
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 442 ----EKPPSSSAPPAPOSLEPVASAHSSASESTSDS-----SSSSPS 483

Oy 733 ESTFKSVSDKK-REKDKPVKSDTKRKSDGSATAKKDNVLLKPSKQOEKVDODREKS 841
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 484 ESESSSSDSEENEPLETAPEPEPTTNKQJDNMLTKYSOPAAPEGRS-----TEP 537

Oy 842 PRSEPPLLKAKEEATKIDSVKSSSQDEKVTGTPRKASHSKASAKDTRRQSPRTRRSR 901
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 538 PRHPESSKSSDSAT-----SQEHSESKDPPKSSSKAPAPAPAPAP-----GKR 583

Oy 902 TVPKTSSQKSQVYR---TRRPSRLKINYLIAREKNERKKKSVDDDESSSKISKY 957
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 564 SQKSPAOEPPQOTVGTGKOPKPKVY-----ASARAGSRISIGEREGLLPYGR 635

Oy 958 EGTETVYKSPKRMKMGDEKLEKTEPKDKIASSTTPAKKIKLNRETGKKIGNAENASTK 1017
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 636 DOTS--KDKPKYKTKG-----RPPAAASN 657

Oy 1018 EPSEKLESTSSKIKOEKYGKAKKRVAGSESSSTLVDT-----STSTGSGPVKRS 1070
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 658 EPKPAVPPSEKHKHKSLLAPASKALSGEPAKDNVEDRTEPHFALVPLTESQGRPHSOS 717

Oy 1071 EKTPTKTVIKTMEYNNNDNTPAEDVIIMIQVQSKMDKDFESEEDVATTPPIQSV 1130
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 718 GSRTGGCAOVAVVQDSKDR-----LPLPL-----RDTLLSPLRLPT 755

Oy 1131 GKPSIILKNVTTKPSATAKYTEKESQPEKLOKLKPEASHELMGLRSSKASASEKGR 1190
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 756 PPOPSLMMKITL-----DLISRIQP-----PGKG---SRQK 785
  
```

```

Oy 1191 AKREHSGSEKDNPDKRKSGAQPDKESTVDRLSEQGHFKTLSSQSKETRTEKHEVNGS 1250
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 786 AEDKPPAGKHHSSSEKRS-----SDSSSKLAKKRKGAEHDC 822

Oy 1231 SKNDETPGRDKVYDSDYSSSKRRDERGELAR-----RKDSPPRKESLSGQ----- 1300
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 823 DNKKLR--LEKEIKSQSSSSSSSHKESKTKTPRSPSSQSKKEMLLPPPVSSSSQKPAKP 880

Oy 1301 --SKLREERDL-----PKKGAESKSN-----SSPPRDK-----PHDK-APYETKRP- 1341
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 881 ALKRSREADTCGODPPPSASTSNHKOSSIPKORVREGKGSNSSSBHGSSGDTANPF 940

Oy 1342 -----CEETKPVDNKSGEREKHAEAARNGKSS 1370
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 941 PVPSLPNGNSKPKQPVAFKQOQDLHNRKAKMKKQKA 978

RESULT
US-08-545-860D-26
: Sequence 26, Application US/08545860D
: Patent No. 6040140
:
: GENERAL INFORMATION:
: APPLICANT: Croce, Carlo
: APPLICANT: Canaan, Eli
: TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
: TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
: TITLE OF INVENTION: Resulting from Chromosome Abnormalities In the All-1 Region
: NUMBER OF SEQUENCES: 94
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
: ADDRESS: No. 6040140ris
: STREET: One Liberty Place, 46th floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/545,860D
: FILING DATE: 07-MAR-1996
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/04436
: FILING DATE: 22-APR-1994
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/10930
: FILING DATE: 09-DEC-1992
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/327,392
: FILING DATE: 19-OCT-1994
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/320,559
: FILING DATE: 11-OCT-1994
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/062,443
: FILING DATE: 14-MAY-1993
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/971,094
: FILING DATE: 30-OCT-1992
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/888,839
: FILING DATE: 27-MAY-1992
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/805,093
: FILING DATE: 11-DEC-1991
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Deluca Esq., Mark
: REGISTRATION NUMBER: 33,229
  
```

```

1 TELEFEEENE/DOCKEN NUMBER: TJU-1262
2
3 TELECOMMUNICATION INFORMATION:
4
5 TELEPHONE: (215) 568-3100
6
7 TELEFAX: (215) 568-3439
8
9 INFORMATION FOR SEQ ID NO: 26:
10
11 SEQUENCE CHARACTERISTICS:
12
13 LENGTH: 1210 amino acids
14
15 type: amino acid
16
17 TOPOLOGY: linear
18
19 MOLECULE TYPE: protein
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
```

Query Match	3.7%	Score 276.5	DB 3	Length 1210
Best Local Similarity	18.3%	pred. No. 3.3e-10		
Matches 227	Conservative 160	Mismatches 428	Indels 423	Gaps 55

QY	231	DITATVTSIVSYSEKSDOPFRDSDUKL-LPAAALTSEHSAGSALTATLAMEEKVGPTSP	289
Db	66	EYKEFLSTKSHTHLDA-----SENRLGKPYPLIPD--KQSS-----IPSSSE	107
QY	290	WNSIFVGOSLHGQL--PTTGPIYRN--AAREGGAPGME-----HSNKLGYLVSPQ-	339
Db	108	HTSV-----HHQSHTPASPGLSVGNISHNPKMAQPRTEPMSLHAKSCG---PPDS	156
QY	340	-----QIRGSESCRSJNRGRHHSERQORSLSLATECFYVPPVPPPLYPPTHLP	395
Db	157	QHLTQDRJGQGEFGSS-----HHKKGRRADG-----DHCAVYT	190
QY	396	PGVPPQFSP--QEPSSQPTAGSYVPPGFPAPANISTAPCSPGVPTAHSN--TMPYT	451
Db	191	DSABERELSPILSPVPP-----LSF-----IHSNQTLPRT	224
QY	452	QAPLLSREEFYREONDKGRESKPEYSGSSYSRSSY--TDSOGLAOHIH-----A	499
Db	225	QG-----SSK--VHSSSNNSKGYCPAKSPKDLAKVHDKENQDSTLYA	265
QY	500	LTLSPSAHNTDLHDHPHPEPEAEASAIIVHMPDLMDIAHAKSSPPYRKRKSRSP	559
Db	266	PAOPPS-----QTFPPPLSPKSVAMQKP--TAYVRPD-----GQDQAP	304
QY	560	PEFRQSPSTKRNVREREKEREYFRNRYKEVP-----PPYDIKAYGRSVDPRDPE	609
Db	305	SE-----SPELKPRLPEDRQOTFEKTDLAKYPAKAKLTKLKMPQSVCOTYSNEYHCYEELL	360
QY	610	KERYREWKRYREMYERYKYGAYAGVAPQPRDSANREDFSPERLLPIRNM--SPTTGR	665
Db	361	KEMTHSW-----PPLTAIHTPSAEPSPKPPFTKDSQHVSVTQNO	402
QY	666	REDYAAQOSHNRNLGON--YPEKLTJRDHNAKNDPKSEKSESENVPGDGKGNKKHRR	723
Db	403	KOYDTSKTSHNSOQGSSMLEDLOJSD-----SEDSSEQTP-----	441
QY	724	KRNRKEKEESESEFLNELL-ETSRKRCGSSGIDETKTDTFLVPSRDATPVYDEPMDA	782
Db	442	-----EKPPSSAPPASAPQSLPEPYASHSSASESTSDS-----DSSSDS	483
QY	783	ESTFYSVSDKDK-REKDKPKVKSDDTKRRKSDGSAITAKNDVILPSKGQOEKYVDGDRKS	841
Db	484	ESESSSSSESENEPLETPAPEPEPTTNKQOLDNWLITKYSOPAAPPECPSPS-----TEP	537
QY	842	PRSEPLTKAAKEEATKIDSVKSSSSQOKDKYVGTGTPKAKHSTKAKXOTRRSQOPRTRRSR	901
Db	538	PRHPDEKSGSDSAT-----SOHSESKODPPKSSSKAPVAPPEAPH---GKR	583
QY	902	TVPKTSQKSOQPVF---TRPRSLRKINTLIAREKNKEREKRSKSDXDFESSMKISKY	957
Db	584	SCQSPSAQOEPQOQTGTKQPKPVK-----ASARAGSRTSLQGEHEPELPIYGRS	635
QY	958	EGTELIVKSPFRKMGDEVKLEKRTPEKDKTASSTTPAKKIKILNREFGKTIAGNAENASTTK	1011
Db	636	DQTS--KQKPYVKTKG-----RPRAAASN	657

[illegible]

```

1      RESULT          7
2      PCT-US94-04496-26
3      ; Sequence 26, Application PC/TUS9404496
4      ; GENERAL INFORMATION:
5      ; APPLICANT: Croce, Carlo
6      ; APPLICANT: Canaanl, Eli
7      ; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
8      ; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
9      ; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1
10     ; NUMBER OF SEQUENCES: 86
11     ; CORRESPONDENCE ADDRESS:
12     ; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
13     ; ADDRESSEE: Norris
14     ; STREET: One Liberty Place, 46th floor
15     ; CITY: Philadelphia
16     ; STATE: Pennsylvania
17     ; COUNTRY: USA
18     ; ZIP: 19103
19     ; COMPUTER READABLE FORM:
20     ; MEDIUM TYPE: Floppy disk
21     ; COMPUTER: IBM PC compatible
22     ; OPERATING SYSTEM: PC-DOS/MS-DOS
23     ; SOFTWARE: Patent Release #1.0, Version #1.25
24     ; CURRENT APPLICATION DATA:
25     ; APPLICATION NUMBER: PCT/US94/04496
26     ; FILING DATE:
27     ; CLASSIFICATION:
28     ; ATTORNEY/AGENT INFORMATION:
29     ; NAME: Deluca Esq., Mark
30     ; REGISTRATION NUMBER: 33,229
31     ; REFERENCE/DOCKET NUMBER: TJU-1242
32     ; TELECOMMUNICATION INFORMATION:
33     ; TELEPHONE: (215) 568-3100
34     ; TELEFAX: (215) 568-3439
35     ; INFORMATION FOR SEQ. ID NO: 26:
36     ; SEQUENCE CHARACTERISTICS:
37     ; LENGTH: 1210 amino acids
38     ; TYPE: amino acid
39     ; TOPOLOGY: linear
40     ; MOLECULE TYPE: protein
41     ; PCT-US94-04496-26

```

Best Local Similarity 18.3%; Pred. No. 3.3e-10;
Matches 227; Conservative 160; Mismatches 428; Indels 423; Gaps 55;

```

QY 231 DITATVSVHSEKSDPFSDNDKL-LPAAALTEHSEKSGASSAITATLMEKVPCTSP 289
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 66 EKEEFLSTKSHTHLDA-----SENRLGPKYPLIPD--KGS-----IPSSP 107
QY 290 WNSIFVQSLHGLI--PTTGPYRIN--AAPGGRPGME-----HSNKLGLVSPQ- 339
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 108 HTSV-----HQSIPTPAGSLVGNISNPKMAQPRPEMPLAKSGC-----PPDS 156
QY 340 ----QIRGERSCYRSINRGHSEKQRTQSPSLPATPCFVVPPLIYPPPHILP 395
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 157 QHUTQDRLGQEGEGSS-----HHKKGDRRADG-----DHCASVT 190
QY 396 PGVPPQFSP--QEPSSQPTAGSVPPPGPPAPANISTACFSPGVPTAHSN--TWPTT 451
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 191 DSAPERLSPILSPVPP-----LSP-----IHSNOQTLPRT 224
QY 452 QAPLLSREFFYREONDKGRSKPPYSGSYRSRY--TDSQGLAHH-----A 499
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 225 QG-----SSK--VHGSSNNKGYCPAKSPKDLAVYHOKETQDSLVA 265
QY 500 LITSPSAHLTLDLHPHPPEEEAARSAMIVMPDLMDIAHARSPPYRYSRGRSP 559
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 266 PAQPPS-----QTFPPPSLPKSVAMQAKP--TAYVRPM-----GDOQAP 304
QY 560 PEPFGOSPTKRVNPREKREYFNRYREV-----PPYDIKAYGKSYDFRDFE 609
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 305 SE-----SPELKPLEDYRQOTFEKTDLKVPAKAKLTILKAMPQSVBOITYSNEVHCVEIL 360
QY 610 KERYREMYREMYEKYKGYAVGAQPPSANEDESPERLPLNIN--SPFTGR 665
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 361 KEMTHSW-----PPPLTAIHPTAEPSPKFPPTKDSQHVSVYQNO 402
QY 666 REDYAAQSHRNRLGNG--YPEKLTSTRDSHNAKDNPKSKESENPVGDGKAKHKKHR 723
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 403 KQYDTSSKTHSNSQGTSMLEDLQSD-----SEDSDEQTP-----441
QY 724 KRNEKEGESESPINPEL-ETSRKCRGSSGIDETDTLFLVPSRDDATPVHDEPMDA 782
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 442 ----EKPPSSAPPAPQSLPEPYASAHSSAESESTDS-----DSSSDS 483
QY 783 ESTFEKSVSDK-RKDKPKVKSDDTKRKSDGSAATAKKNVILKPSGPOEKVDREKS 841
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 484 ESESSSSDSENEPLETPAPEPEPTTNKQOLDNMLTKVSOAPAPPGPRS-----TEP 537
QY 842 PRSEPLAKKKEATKIDSVKPSSSSOKDEKVTGTPRKAKSKAKDPRROQPTRRSKR 901
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 538 PRHPESKGSDDAT-----SOEHSKSDPPKSSSKAPRAPPEAPH--GKR 583
QY 902 TVPTSSQKQPVY--TRPRSLRKINYLIAREKNEREKRSKVDKDFESSMKISKV 957
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 584 SCQKSPAQOEPPOQOTGTQPKPKPVK-----ASARAGSRKSLQGERPGLLPGR 635
QY 958 EGTETVPSPKRMGEVEKLETPERKOKIASSTTPAKKIKLNETGKKIGMAENASTTK 1017
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 636 DQNS--KDKRKVKTKG-----RPRRAASN 657
QY 1018 EPEKLESTSSKIQEYKVGKAKKAVAGSEGSSTLYDT-----STSGGSPYRKS 1070
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 658 EPRPAVPPSEKKHKSSLPAPSKALSGPEPAKNVEDRTPENHALVPLTESQPPHSGS 717
QY 1071 EEKTDTRVYIKTMEENNNNTAPAEVYIMIOVQSKMKWDDESEDEVDYKTTQIOSV 1130
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 718 GSRTSGCRAVVVOEDSRKR-----LPLPL-----RDKLLSLPLRDT 755
QY 1131 GKPSIITKNTYTPSATAKYTEKESEQPEKLOLKPKEASHLMQHELRSKSGASSEKGR 1190
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 756 PPQSLVWKTTL-----DLSTRIPQ-----PKG--SRQK 785
QY 1191 AKDREHSGSKNDPKKSGAQDPKESTVDRLEQGHFKTILSOSKKTTRTSEKHESVRGS 1250
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

Db 786 AEDKOPPAKKHSEKRS-----SDSSSLAKKRRKGEAERDC 822
QY 1251 SNKDFTEGROKKVYDSDSSKRRDERGLAR-----RKDSPPKGESLSGOK---- 1300
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 823 DNKKIR--LEKEITSQSSSSSSSHKESKTKTPSPSSQSKKMLPPPPVSSSQKAPK 880
QY 1301 --SKLREERDL--PKGAESKRSN--SSPPDK-----PHDK-APYETKRP- 1341
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 881 ALKRSREADTCGDDPPKSASTKSNHKDSSIPKQRVKEGKSRSSSEHKSSGDTANPF 940
QY 1342 -----CEETPKVDKNGKEREKHAAPARNKSS 1370
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 941 PVPSLPNGNSKPKPYKQDFKQADLHMRKAKMKQKA 978

```

```

RESULT 8
US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/741,940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
; US-07-741-940-7

```

Query Match 3.7%; Score 274; DB 1; Length 2842;
Best Local Similarity 18.5%; Pred. No. 1.5e-09;
Matches 282; Conservative 188; Mismatches 612; Indels 446; Gaps 61;

OY	173	FRIVYSPCSDTKRAGSCDSOSGLSLPLAPATISLSLNSQ--SLAPVYSGNPPSAPVY	230
Db	1194	FSFKSSSGSSGSSKTEHMSSESSENTS---TPSSNAKRQNDLHPSASQSGOPORA-----	1245
OY	231	DITATVATISVHSEKS-----DGPF---RDSDNKLPLPAA-----LTSEHSKASSIA	274
Db	1246	---ATCVKSSINQETIOTYCEVEDPTICFNSCSSLSSLSAEDEICNOTTOEADBSANLQ	1302
OY	275	ITALMEERG-----VPTGSP-----WNSITVGOSLHGOILPTTG---PVR	312
Db	1303	IAELKEKIGTRESADPVEVPAVSOHPRTKSSRLQSSLSSESARHKAAVEFSGAKSPSK	1362
OY	313	INAPRGGRPGMEHMSNKLGLVLSPOQILRGERSOCYRSINGRHHSEKSOQSPSLPA	372
Db	1363	SGAOTP-----KSPBEHYQETPLMFSTCTSVSSLSIDFESNIASSVQS	1406
OY	373	TPCEVPV-----PPPLYPPPPHITPLRPGVPQPSF---QFPSSOPTAGVSP	420
Db	1407	EPCSGMVSGLISPDLPDSGOTMPPSRSKTPE--PPQTAQTKREVPRKNKAPTAREKS	1464
OY	421	PRG-----FPPAPA--NISTACFSGVYTAHNTMPTQOAPLSREFFYR--	463
Db	1465	GPKOAAVAAVORYOVLPLDDTLHLHATBESTPBGFCSSLSLALSIDEPFIQDKVELIM	1524
OY	464	---BONDGRKREKPPYSGSSYSRSSYTDSSOGIAOHIAHLTLPASAHTLDLHHPHP	520
Db	1525	PRVQENDNGNTE-----SQPKESNENQEKEMKIDT-----EKLILDSDDDD	1570
OY	521	EBEARSAMIVHMDLMDIAHARSRPYRKYSRSRSPDEFR--GOSPTRKNVPREK	577
Db	1571	IEI-LEECILSAMP-----TKSSRKAKPAQTAKLTPAPARRPOLPYKILLPSQNR	1622
OY	578	EREYFNRYREVPPEYDIKAY---GGSYPRDPEFEKERYREMERKYYREMYEKYYGAVG	634
Db	1623	LQP---QKHVSTPEDDMDPRVYCYEVPINSTATSLIDL-----IESPPNELAAG	1671
OY	635	AQPRPSANREDFSEBRLPLNIRNSPFTTRGRREDYAAAGSHRRNRNLGNYPEPKLSTRSH	694
Db	1672	EGVGGAQSGEFKEKROTIP-----TEGRSTDEAG-----GKTSSVTIPELDND	1715
OY	695	NAKON-----PKSKEKE-----SENVPG-----DGKGNKHKHR	723
Db	1716	KAEGBDILAECSINAMPKGSHKPRFYKKIMDOVOASASSAPAKNOLDKKKPPJSPV	1775
OY	724	K---RNEKEGGESESFLNPelletTSRKCRRSSSIDETKIDTTLFVLRSDA---772	
Db	1776	KPIQNTYERIRRYAKNADSKNNLNLAEBVEPDNDKSKQKYLKNNSDFNDKLPLNNDRVRG	1835
OY	773	-----TPVDEPM-----DAESITFKSVSDKREKDKPRVYASDKTKRKSDDGA	816
Db	1836	SFAPDSPHHYPIRGTFYCFPSRNDSSLDF---DDDDVDLSREKAEJLAKKEKESEEA	1891
OY	817	TAKDNVLPKPKGQOEVDVDREKSPRSPEPLKAKAEATKIDSYK--PSSSSQDKEK--	872
Db	1892	KVTSITELTSSQASANKTOALIAQIPINROGPKITLOKSTFFQSSKIDPDKCAATDELQ	1951
OY	873	---VTGTP-----RKAHSAKANDTR--ROGOPRTKRSKRTVPTS	907
Db	1952	NFAIENTPVCFSHNSSLSLSDIDQENNNKENEPKETEPDPSGCBPMQASGVAPKSF	2011
OY	908	SOKQOPVATRPRRLKRIYN---LI-----AREKNEREK-----	938
Db	2012	HVEDTPVCFSHNSSLSLSDISEDDLLQECISAMPKKKPRLRKLDGDNKHSPRNMGIL	2071
OY	939	-----RKSVDK	945
Db	2072	GEDTLTLDKIQRPDSBHGSLPSENPDMKAIQEGANSIVSLHQAAAAACLSRQASDS	2131
OY	946	DE-----ESSSMKISYEGEILYIPBPKKMGVDEKLERTPEKIKI	987
Db	2132	DSIUSLKGISLSPFHLTPQOEKEPFTSKGRRLLPKPEKSTLE--TKKIESESKGK-2188	

Qy	988	ASSTTPAKKILKINRENGKATGKNENNASTTRKEPEKLESTSSKIKOEKYGAKKRRVASE	1047
Db	2189	-----GKKRYKSLITGKVRNSEISGOMOP---LOAMPIS---RGRTMHPeVR	2236
Qy	1048	GSSTLYDVYSTSTSGSPVR-----KSEKTYTKRVITKMEYVANNMDTAAEDVTIMI-	11020
Db	2237	NSSSS-----TSYPKKKGPPLTKPASKSPSGCGATISPRGAPVSKSELSPVARTQ9IG	22922
Qy	1103	---OVPOSKMDKDFESEEEDVTTOPPIOsvGKPS-----SITKN	1139
Db	2293	GSSKAPRSRSGSRSTSPRPAQPLSRPISPGNNSISPGRNGISPBNKLSOLPRTSBSPT	23522
Qy	1140	VTTKPSATK--YTE-----KSEDOPEKLOKLEPKAEHELMOHELRSKGSASS	1186
Db	2353	ASTSSSGGKMSTSPEROMSQONLTKOYGLSKMNASIPRSHASAKGLGNOMNNGKANUK	2412
Qy	1187	EKGRAKDREHSGSEKDNPK---KSGAOPDKESTYDR--LSOQHFTKLSOSK--	1236
Db	2413	VELRMSSTKSSGSEDSRSPVLVQOSTFIKEAPSPTLRRKLEBSASFESTLSPSRAS	24722
Qy	1237	ETPTSEK-----HESVRGSSNMDFPNGRKKKVDYDSRYSKSKRDEGE	1281
Db	2473	PTRSQAQPVLSPLSLPMSLSTHSSVQAGGKRLPNLSPTIEIN--DGRPKRHD----	2526
Qy	1282	LAR-RKDSPPRGKESLEG--OKSKLREERDLPKKAESKSKSSPPDKKPHDHAPEYT	1338
Db	2527	IARSHSESPRLPINRSGTWKREHKSHSSSLPVRVTRWRTGSSSSILASSESE-----	2581
Qy	1339	KRPCEETKPVDKNGSKREKHAALRNG	1366
Db	2582	KAKSEDEKHNVISIGTKOSKENOVASKG	2609

RESULT 9
 US-08-289-548A-7
 Sequence 7, Application US/08289548A
 Patent No. 5648212
 GENERAL INFORMATION:
 APPLICANT: ALBERTSEN, HANS
 APPLICANT: ANAND, RAKESH
 APPLICANT: CARLSON, MARY
 APPLICANT: GRODEN, JOANNA
 APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSLYN, GEOF
 APPLICANT: KINZLER, KENNETH
 APPLICANT: NAKAMURA, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THILYERS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 102
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner & Allegretti, LTD
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20001-4598
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/289,548A
 FILING DATE: 12-AUG-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107.46943
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2842 amino acids
 TYPE: amino acid
 STRADEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: APC
 US-03-289-548A-7

Query Match 3.7%; Score 274; DB 1; Length 2842;
 Best Local Similarity 18.5%; Pred. No. 1.5e-09;
 Matches 282; Conservative 188; Mismatches 612; Indels 446; Gaps 61;

173 FRYTSPYSCDTKAGSCSDGTLRLPAPISISLTNO--SLAPVYSGNPPSAPAPVP 230
 1194 FFSKSSSGQSSKTEHMSSESTNTS--TPSNARQQLHPHSSAQSHSGQPKA----- 1245
 221 DITATVSISSVSEKS-----DGF--RSDNKLPLAA-----LTSEHSGASSIA 274
 1246 --ATCKVSSINQETIQTYCEDPTICFSCSSLSLSSADEIGCNCTQTEADANTLQ 1302
 275 ITALMEEG-----VGTSP-----WNSIFVGOSLHGQLIPTTG--PVR 312
 1303 IAEIKETIGTSAEDPVEVAVASHPRTKSSRLGSSLSLSSAHKAVERSGAKSKS 1362
 313 INAAPGGGPRGWEHNSKGLVSPPOQIRGERSCYRSINGRHNSERSQTOGSPSLPA 372
 1363 SGAQTP-----KSPENHYQETPLMFSCRSTSVSLDSFESSRISAVQS 1406
 373 TPCEFPV-----PPPLXPPPHTLPLPGVPPQFSP--QFSSQPPPTAGIYVP 420
 1407 EPCSGWVGIIISPSDLPSPQTMPPSRSKTTP--PPQTAQTKREVKNAPTEKRES 1464
 421 PRG-----FPAPA--NISTACFSPGVTAHNTMPTQAPLISREEFYR-- 463
 1465 GPKQAAVAVAVQVQVLPDALTLHLFATESIPDFGSCSSLSLSSLDPEFLQKDELKIM 1524
 464 --BONDKRESKPPYSGSSYSRSSYTDSSQLAOHIALTLSPSAHTLDDLHDHPRP 520
 1525 PPVQENDNGNETE-----SEQPKESNENQKEAEKTIDS-----EKDLDDSDDD 1570
 521 EEAERASMIYHMDLMDIAHAKRSPPYRRYRSKRSRSPPEFR--GOSPTRKRVNPREK 577
 1571 IEI-LEECIIISAMP-----TKSSRKAKKPAQTAASKLPPVARKPSQLPYKILPSQNR 1622
 578 EREYFNRYREVPPIYDIKAVY--GRSYVDFDPFEKERYREMERKRYRMVEKYYGVAVG 634
 1623 LQP--QKIVSTTPGDDMRYVCVECTPIINFSTATSLSLT-----IESPPNELAAG 1671
 635 AQPRESANREDFSPERLLPLNIRNSPTTRGRREDYAAQOSHNRNMLGNYPEKLTSTRSH 694
 1672 EGVGGAGQSGFEKRDITP-----TEGSTDEAOG-----GKTSSVTIIELDON 1715
 695 NAKDN-----PKSKEK-----SENVPG--DGKGNKHKHR 723
 1716 KAEGBDILAEICINSAMPKSHKPPRVKKINDVOOASASSAPKKNOLDGKKKPTSPV 1775
 724 K-----RNRNEKGESESEFLNPELLETSRKCGSSGIDETKDTLFLVPSRDA--- 772
 1776 KPPOINTETFRVRKNKADSKNNLNAERYSDNKSCKONLNNKSKDENDKLPNNEDRYRG 1835
 773 -----TPVDEPW-----DAESITPKSVSDKDKREKDKPKVSDTKTKRSSDGA 816
 1836 SFARSPHHYPIEGTPICFSRNDSLSLD-----DDDDVDLSREKAKLRAKAKKEKEEA 1891
 817 TAKDNVILPKSPGQOEKVDGDKREKSPNSEPPLKAKKEATKIDSVK--PSSSSQKDEK-- 872

1892 KYTSHTTELTISNOQSANKTOAIKOPINRGOPKPILOKOSTEPPOSSKIDPDGCAATDEKQ 1951
 873 ---VTGTP-----RKASKSADOTR---ROGOPTRRSKRTVPKTS 907
 1952 NFAIENTPVCSHNSLSLSDIDOEENNNKENEPIKETEPPDSQGEPSKPAAGTAPASF 2011
 908 SOKSQVTRTRPRSLRKINY-----LI-----AREKNEREK----- 938
 2012 HVEDPVCFSNRSSLSLSDSEDDLQECISAMPKKKKPSRLKGDNEKHSPRMGGIL 2071
 939 -----RKKSVDK 945
 2072 GEDTLDLKDIDRPDSEHGLSPDSENFDMKAIQEGANSIVSLHQAAAAACLSPDASSDS 2131
 946 DF-----ESSMKISKVGTETIVKPSPKRMGDVEKLETPPEKDI 987
 2132 DSLSLKSGISLSPFHLTPDOEEKPFSNKNCPRLKKGKSTLE--TKKIESEKGIK- 2188
 988 ASSTTPAKKIKLNRETGKIGNAENASTTKPESEKLESTSSKIKQEKYKAKRRVAGSE 1047
 2189 ---GKKVYKSLITGKVRNSSEISGOMKP--LQANMPSIS---RGRTHIHIPGVR 2236
 1048 GSSSTLVDTYSTSSVSGSPVR---KSEKTDKRTVYIKTMEYVNDNTAFAEDVIIMI- 1102
 2237 NSSSS---TSPVSKKGPLTPASKSPSEGTATTSPRGAKPSYKSELSPVAROTSQIG 2292
 1103 ---GVPOGKWDKDFESEEDVKTPTPIQSVGKPS-----SIITKN 1139
 2293 GSKKAPRSRSGRSDTSPSPAPQOPLSRPIQSPGRNISIPGRNGISPPNKLSQLPRSSPST 2352
 1140 VTKPSATAK--YIE-----KESEDEKLOKLPEKASHLMOHELRSKSGASS 1186
 2353 ASTKSSGSGKMSYSPGROMSQONLTQOTGLSKNASSIPRESASAKGLNOMNNGANKK 2412
 1187 -EKGRAKREHSGSEKDPKDP---RKSQAQPDKESTYDR--LSPOGHFKTLSSQSK--- 1236
 2413 VELSRMSTKSSGSESDSEREVLYROSTFIKEADSPILRKRLLESASFEESLSPSRAS 2472
 1237 ETRTSEK-----HESVRGSSNNKDFTPGRDKVYDSDRDSYSSSKRRDEGE 1281
 2473 PTRSGAQTPVLSPPLDMSLSTHSSVQAGWGRKLPNLSPTIEYN--DGRPAKRHD----- 2526
 1282 LAR-RKDSPPRGKESLSC--OKSKLREERDLPKKAESKSSNSPPRDKKPHDKAPYET 1338
 2527 IARSHSESPLRPIRNSGTMRKREHSHSSSLPRVSTWMTGTSSSILSASSESE----- 2581
 1339 KRPCETKPVDNKSGEKEREKHAAPRNG 1366
 2582 KAKSEDEKHVNSISGTKOSKENQVSAKG 2609

RESULT 10
 US-08-452-654-7
 Sequence 7, Application US/08452654
 Patent No. 5691454
 GENERAL INFORMATION:
 APPLICANT: ALBERTSEN, HANS
 APPLICANT: ANAND, RAKESH
 APPLICANT: CARLSON, MARY
 APPLICANT: GRODEN, JOANNA
 APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSLYN, GEOFF
 APPLICANT: KINZLER, KENNETH
 APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THALIVERTS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 94
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Banner, Birch, Mckie & Beckett
 STREET: 1001 G Street, NW
 CITY: Washington

STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,654
FILING DATE: 25-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-452-654-7

Query Match 3.7%; Score 274; DB 1; Length 2842;
Best Local Similarity 18.5%; Pred. No. 1.5e-09;
Matches 282; Conservative 188; Mismatches 612; Indels 446; Gaps 61;

173 FRYVSPGSDPTKTAGSCSDSGTSLRLPAPSTISLTNSQ--SSLAPVSGNPPSAPAPVP 230
1194 FFSKSSSGSSGSKTEHMSSESENTS--TPSSNAKRQOLHPSSAQSSGQPKA----- 1245
231 DITAVSVISVSEKS-----DGF-----RSDNKLPLPAA-----LTSESKASSIA 274
1246 ---ATCKVSSINQETIQTVCVEDTPICFSCSSLSLSAEDICGNTQADSANLTQ 1302
275 ITALMEKG-----VPGTSP-----WNSIFVGSLHGLIPTTG--PYR 312
1303 IAEIEKIGTRAEDEPVSEVPAVSQHPRTKSSRLGGSSLSLSAHHKAVESSGAKSPK 1362
313 INAAPGGGPRGWEHNSKGLVLPSPQOIRGERSCYRSINGRHHSERQORTQSPSLPA 372
1363 SGAQRP-----KSPPEHYVQETPLMFSSRCTSVSSLDSPESHSSIASVGS 1406
373 TPCEFPV-----PPPLYPPLPHLPLPPVPPQFSP--QFESSOPTAGYSVP 420
1407 EPCSGVGIISPSDLPSPGQTMPPSRSKTTP--PPQTAQTKREVPPKNKAPTAEKRS 1464
421 PPG-----PPAPA--NISTACFSPGVPTAHSNMPPTQAPLREEFYR-- 463
1465 GPKQAANAAYOVOLVDADTLHMFATESPTDGFSCSSLSALSIDEFTIOKVDLRIM 1524
464 ---EONDGREGSKFPYSGSSYSRSSYTDSSQGLAQHIALHLLSPAAHTLIDLHPHP 520
1525 PVOENDGNENETE-----SEQPKESNENQEKAEKIDS-----EKDLDDSDDD 1570
521 EEAERASMIYVMDLMIHAHRSRSPYRRKRSRSPREPR---GQSPTKRNVPREK 577
1571 IET-IEECTISAMP-----TKSSRAKAKPAQTASKLPVPYAKRPSQLPVYKLLPSQNR 1622
578 EREYFNRYREVPPIYDIKAYV--GRSVDFRDPFEKERYREWERKYREMYERYKGYAVG 634

1623 LQP--QKHVSFTPGDDMPRVYCVVESTPINFSTATSLDLT-----IESPPNELAAG 1671
635 AQPSPANKREDESPERLLPLNIRNSPFTGRREDYAAQSHRRNMLGNYPEKLSRDSH 694
1672 EGVRCGAQSGEKEKRDITP-----TEGRSTDEAOG-----GKTSSVTIPELDN 1715
695 NAKDN-----PKSKEKE-----SENVPG--DGGNKHKKHR 723
1716 KAEEDILAEICINSAMPKGSHPKPPRVKIMQVOQASASSAPKNQDLGKKKPTSPV 1775
724 K-----RRNEEGEESFLNPELLETSRKCGSSGIDTQTDTLFVLPSRDA--- 772
1776 KIPONTERTVYRKNAADSKNNLNAERVSDMKDKQVLLKNNKSDFNKLPNNEDRVAG 1835
773 -----TPVRDEPM-----DAESTIFKSVSDKOKREKDKPKVSDKTKRKS DSGSA 816
1836 SFAPDSPHHYPTIEGTPYCFSRNDSLSLDF-----DDDDVDLSREKALRAKENKESPA 1891
817 TAKKNVILKPSKGPQEKVDGDEKSPREPLKKAKEATKIDSVK--PSSSQDEK-- 872
1892 KVTSHTELTLSNQASANKQALAKOPINRGQPKPILQKOSTPQSSKIDIPDRGAATDEKQ 1951
873 ---VTGTP-----RKAHKSADTR--ROSOPTRRSKRTVPKTS 907
1952 NFAIENTPVCFSHNSLSLSDIDQENNNKENEPKETEPPOSGEPKSPQASGYAPKSF 2011
908 SOKSQPVTRRRPSLRKINY-----LI-----AREKNERK----- 938
2012 HVEDTPVCFSRNSSLISLIDSEDDLQECISSAMPKKKKPRLKGDNEKHSPPRMNGITL 2071
939 -----RKKSVDK 945
2072 GEDLTLDLKDIOQRPDSEHGLSPDSENFPMKALQEGANSTIVSSLHOAAAAACLSROASSS 2131
946 DE-----ESSMKISKVEGTEIYKPSPKRMGADVKEKLETPPEKDKI 987
2132 DSILSLKSGISLSPFHLTPDDEKRPFSNKNKPRILKPEKKTLE--TKKISESKGIR- 2188
988 ASSTPPAKKIKINRETGKKIGNAENASTTKEPSEKLESTSSKIKOEKVKAKARRVAGSE 1047
2189 -----GKKVYKSLITGKVRNSSEISGOMKP---LQANMPIS--RGRTHIHIPGVR 2236
1048 GSSSLVDVDTSSSTGGSPVR--KSEKTDTKRYIKTMEIYNNDNAPADVIYIM- 1102
2237 NSSSS---TSPVSKKGPPLKTPASKSPSEGTATTPRGAKPSVKSSELSPVARQTSQIG 2292
1103 ---QVPSKMDKDDPESEEDVKTTOPIQSVGKPS-----SIKX 1139
2293 GSSKAPSRGSDSTPSRAQOPLSRPIOSPGRNISTPGRNISTPKNLSQLPRTSSPST 2352
1140 VTKPSATAK--YTE-----KESEQPEKLOKLPEKASHLMOHELRSKGSASS 1186
2353 ASTKSSGSKMGYTSPPROMSOONLTKGTGLSKNNSIIPRESASAKGLNONGNANK 2412
1187 -EKGRAKREHSGSEKDNNDK--RKSQAOPDKESTYDR--LSQDGHKTTLSQSK-- 1236
2413 VELSRASSTKSSGSDSRSEPVLRQSTFIKEABSPTLRRKLEBSASESLSPSSRPAS 2472
1237 ERTSEK-----HEVRGSSNKDFTPPGRDKYVDSDRVSSSKRDERGE 1281
2473 PTRSQAQTPVLSPLPDMKLSHSSVQAGWKKLPNLNLSPTIEYN--DORPAKRHD-- 2526
1282 IAR-RKDSPPRKESLIG--QKSKLREERDLPPKGAESKSSNSPPROKKPHDKAPYET 1338
2527 IARSHSESPSRPLPIRNSGTRKREHSHSSSLPRVSTWRTGSSSITLSASSSESE-- 2581
1339 KRPEETKPVDNKSGKEREKHAALRANG 1366
2582 KAKSEDEKHVANSISGTQSKQENQVSAKG 2609

RESULT 11


```

OY 1187 -EKRAADREHSGSEKONPDK-----RKSQAQPDKESTYDR--LSQGHFKTILTSQSK--- 12366
Db 2414 VELSRMSSTKSSGSESDSRSPVLVKNQSTFIKEAPSPTLRRKLKLEASAFESILSPSSRPAS 2473
OY 1237 ETFTSEK-----HESYRGSSNKDFPTPGRDKKVDYSDRPYSSSKRRDERGE 1281
Db 2474 PTRSOAQTPVLSPSLPDMSLSTHSSVQAGWKRRLPNLSPTLEYN--DGRPAKRND----- 2527
OY 1282 LAR-RKDSPPRGKESLUG--QKSKLREERDLPKKAESKSKSSPPRDKKPHDKAPYET 1338
Db 2528 IARSHSESPRLRLINRGCTKRRSHKSSSLPVSFTWRTGSSSILASSESE----- 2582
OY 1339 KRPEETKPYDKNSGKREKREHNAEAANG 1366
Db 2583 KASEDEKHWNSISGTIKQKRENOVSAGK 2610

RESULT 12
US-08-452-655B-7
Sequence 7, Application US/08452655B
Patent No. 5783666
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1001 G Street, NW
City: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,655B
FILING DATE: 25-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES

```

[illegible]

```

Db 2073 GEDTLIDLKDIOQRPDSEHGLSPDSENDMAKIQEGANSIYVSLHQAAAACLSRQASSDS 2132
QY 946 DF-----ESSSMKISKVEGIEIYKPSKRMKEGVLEKTPPEKDKT 987
Db 2133 DLSLSKSGISLSPHLPDQOEKPTSPKGRILKPGKSTLE--TKKIESESOKIK- 2189
QY 988 ASSTTPAKKIKLRENGKTKIGNAENASTKEPSEKLESTSSKIKOEKVKAKRKVAGSE 1047
Db 2190 -----GKKVKYKSLITGVKANSSEISGOMKOP---LOANPSPIS---KRTMHIHIGVR 2237
QY 1048 GSSSTLVDTSTSTSGSPVR---KSEKTDKRTVYIKTMEEYNDNTPAPADVIIMI- 1102
Db 2238 NSSSS---TSPVSKKGPPLKTPASKSPSEGGATATSPRGAKPSVKSELSPVAROTSQIG 2293
QY 1103 ---QVQSKKDKDFESEEDVKTPTDIOGVGRS-----SITKN 1139
Db 2294 GSSKAPSRSGSRDSTPSRPAQPLSRPIQSPGRNSISPGNGISPPNKLSQLPRTSPST 2353
QY 1140 VTKRPSATAK--YTE-----KESQPEKLOKLPKEASHELMQHELRSKGSASS 1186
Db 2354 ASKSSSGSGMSTTSPGRQMSQONLTKOTGLSKNASSIPRSEASKGLNOMNGNANKK 2413
QY 1187 -EKGRADREHSGSEKDNPKD---RKSGAPDKESTVDR--LSEOGHFTLSQSSK-- 1236
Db 2414 VELSRMSSTKSGSESDRSERPLVROSTFIKEAPSTLRKLEESASFESLSPSRPAS 2473
QY 1237 ETTTSEK-----HESVSGSNKOPTPERDKKVDYDSDYSSSKRRDERGE 1281
Db 2474 PTTSQOQTPVLSPLDPMLSLTHSSVQAGWRKLPPLISPTIEYN--DGPARRHD--- 2527
QY 1282 LAR-RKDSPPRGKESLSG--QKSLREERDLPKKGAEKSKNSPPDKKPHHKAPYET 1338
Db 2528 IASHHSPSPRLILNSGTGKREHSHSSSLPVSTWRTKRGSSSSITLSASESEF----- 2582
QY 1339 KRPCETKPYDKNSGKEREKHAAEARNG 1366
Db 2583 KAKSEDEKHVNSISGTQSKENQVSAGK 2610

```

RESULT 13
US-08-450-582-2
; Sequence 2, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MART
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THILVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Wilcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,582
; FILING DATE:

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/452,655
; FILING DATE: 25-MAY-1995
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107,49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amlng acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-450-582-2

```

```

Query Match 3.7%; Score 274; DB 3; Length 2843;
Best local Similarity 18.5%; Pred. No. 1,5e-09;
Matches 282; Conservative 188; Mismatches 612; Indels 446; Gaps 61;

```

```

QY 173 FRYTSPYSCDRTTASGSDGTLRLPAPISLTSNQ--SSLAPVSGNPSAPAPVP 230
Db 1195 FSPSKSSSGSSKTEHMSSESTMS--TPSSNAKRONQLHPSSAOSGQPOKA----- 1246
QY 231 DITATYISIVHSEKS-----DGPFF--RDSNKLPLPAA-----LTSEHSGKASSTA 274
Db 1247 --ATCKVSSINQETIQTVCEDTPICFSCSSLSLSAEDIEGQNTQOEDSANTLQ 1303
QY 275 ITALMEEK-----VPGTSP-----WNSIFVGSLHGLIPTTG--PVR 312
Db 1304 IAIKEIKIGTASADRYSEVPANVQHPRTKSSRLQSSLSLSEBARHKAIVFSGAGASPSK 1363
QY 313 INARPGGGRGWEHSHNKLGLVSPQOIRGERSCTRSINGRHHSNQRTQSPSLPA 372
Db 1364 SGAQTP-----KSPPEHYVQETPLMFSTRCTSVSSLSDFESRSIASSVOS 1407
QY 373 TPCGFVPY-----PPPELPPPTPLPVGVPPOPS--QPPSQPTAGSVY 420
Db 1408 EPCSGMVSGLTSPSLDPSFGOTMPPSRKTPP--PPQTAQTKREVPKKAFTAEKRES 1465
QY 421 PPG-----FPPAPA--NISTACFSPGVPTAHSNTMPTTQAPLLSREFFYR-- 463
Db 1466 GPRQAAVNAVQRYOVLPRADTLHFATESTPDGFCSSSLASLSDEPIQDVELRIIM 1525
QY 464 ----EONDKGRSKFPYSGSSYSSTSDSOGLAQIHIALTLPSAANTLIDLADHPHR 520
Db 1526 PRYQENDNGNETE----SEQKRESNDEKAEKIDS-----EKDLDDSDDDD 1571
QY 521 EEAARSAMTVHPMDLMDIHAHRSRSPRYRSRSRSPPEFR---GQSPTKRVPREEK 577
Db 1572 TEL-LEECITISAMP-----TKSSRKAKKPAOTAKSLPVPVARKPSQLPVYLLPSQNR 1623
QY 578 EREYFNRYREVPPIYDIKAYY---GRSVDFRDPFEKERYREWERKYKGYAVG 634
Db 1624 LQP--QKHVSFTPGDMPRVYCVGEGTPIINSTATSLSDLT-----TESPNEIAAG 1672
QY 635 AQPSPANREDESPEKLLPINRNSPPTKGRREDYAGOSHRRNRNIGATPELSTROSH 694
Db 1673 EGVRGAGQSGEFKRDTP-----TEGRSTDEAG-----GKTSVVTIPELDN 1716
QY 695 NAKDN-----PKSKEKE-----SENVPG--DGKGNKHKHHR 723
Db 1717 KAEGLDLAECINSAMPKGSKHKPFYKKTIMDOYQOASASSAPNKNQDLGKAKKKTSPV 1776

```

```

0Y 724 K-----RRNEKGESESSFLPNLLTFSKRCRSSSIDETKTDLPVPSRDA---- 772
Db 1777 KPIPDNTEYTRVRKKNADSNNMLNARVSDNKNDSKKOVLKNNSKDFNDKLPNNDRVGR 1836
0Y 773 -----TPRDEPM-----DAESITFKSVSDKDRKDKPRKVSCKTKRKRSQDGA 816
Db 1837 SFADSPHHHTPIEGTPIYCFSRNDSLSSIDF-----DDDDVDLSREKAELEKRAKEKESEA 18922
0Y 817 TAKDNVLPKSKGQOEKVDGDRKESRSPSEPLKKAKEEATKIDSVK--PSSSSQDEK-- 872
Db 1893 KVTSHTELTINQOASANTQAIKAPINRGQPKPILOKOSTFPOSSKDIPIIDRCATDEKLO 19522
0Y 873 ---VTGR-----KRAHSAADTR---ROQOPTRRSKTRVPTS 907
Db 1953 NFALETPPVCFSHNSLSSIDIDQENNNKENEPIKETEPPDOGEPSKPOASGYAPKSF 20122
0Y 908 SOKSQPRTRRPRLKINX-----LI-----ARENNEREK----- 938
Db 2013 HVEITPCFSRNSLSLSDISEDDILOECISSAMPKKKKPRILKGDNEKHPRMGIL 20722
0Y 939 -----RRKSVDK 945
Db 2073 GEDLTLDLKIDQRPDEHGLSPSENPDMWKAIOEGANSIVSSILHOAAAAACLSRQASDS 21322
0Y 946 DF-----ESSSMKISKEVEGEIYKPPPKKMGDEVKLETRTEKTKI 987
Db 2133 DSIILSKGISLSPFHLTPDOEKEPFTSNKKGRIILKPEKSTLE--TKKISESGKTK- 2189
0Y 988 ASSTPAKKIKLNETEKTKIGNAENMSTTKREPEKLESTSKIKOEKVGAKKRVAGSE 1047
Db 2190 -----GKKRYKSLITCKVANSSEIGOMKOP---LOAMPSTIS---RGRTMHITPGVR 22372
0Y 1048 GSSSTLVDTYSTSGSGSPVR---KSEKTDTKRTVIRTMEYNNNDTAPAEVYIIM- 11022
Db 2238 NSSSS-----TSVYSKKGPIKTPASKSPSBOGATATSPRGAPVYSKSELSPARQTSIG 22932
0Y 1103 ---QVPOSKMDKDFEEDVDYTTQPIQSVKPS-----SITKN 11392
Db 2294 GSSKAPRSRSGRSTSPSPAQOPLSRPIOSPGRNLSIPGRNLSIPPNKLISQLPRTSSPT 23532
0Y 1140 VTTSPASATK--YTE-----KESQPEKTLKLEKASHELMHQHELSSGSASS 11862
Db 2354 ASTSSSGGKMKSTSPGRQMSOONLTQGLSKNASSITPSESASKGLQMMNGGANKK 24132
0Y 1187 -EKRAADREHSGSEKNDPDK---KSGAQPDKESTYDR--LSEOGHFTKTLSSG-- 12362
Db 2414 VELRMSSTSSGSEDSRSPVLYVQSTFIKAPPTLIRKLEESASGESLSPSRAS 24732
0Y 1237 ETRTSEK-----HESYRGSSNKKDFTRGDKKVDYDSRYSSSKRRDERGE 12812
Db 2474 PTRSOAOTPVLPSPILPMSLSTHSSVQAGWKRKLPNLSPTIEYN--DGRPAKRHD--- 25272
0Y 1282 LAR-RKOSPPRGKSLSG--QSKLEERLPLPKKGAESKSNSSPPRODKPHDKAPET 13382
Db 2528 IARSHSPSPPLPINRGWTMKRSHKSSSLPVSPTWARTGSSSITLSASESSE----- 25822
0Y 1339 KRPEETKPYDNKSGKEREKHAAEARNG 1366
Db 2583 KAKSEDEKHNVISGTKQSKDENOVSAKG 2610

RESULT 14
US-08-450-582-7
; Sequence 7, Application US/08450582
; Patent No. 6114124
;
; GENERAL INFORMATION:
;
; APPLICANT: ALBERTSEN, HANS
;
; APPLICANT: ANAND, RAKESH
;
; APPLICANT: CARLSON, MARY
;
; APPLICANT: GRODEN, JOANNA
;
; APPLICANT: HEDGE, PHILIP J.
;
; APPLICANT: JOSIYU, GEOFF
;
; APPLICANT: KINZLER, KENNETH
;

```

APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450,582
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/452,655
FILING DATE: 25-MAY-1995
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
US-08-450-582-7

[illegible]

Qy	421	PRG-----	FRPAPA--N1STCSPBGPVTAHNSMMPQOAPLISREEF--	463
Db	1466	GPKOAAVNAAYORVQVLPDLDLTLHATSTPDGFCSSSLSAISLDEPFIQKVLEIRIM	1525	
Qy	464	---EONDKGRESFEFYSGSSYSSTYDSOOLAHIALTLSPSAHTLIDLHDHPH	520	
Db	1526	PRVGENONGMTE-----SQOPRESHENEGEKAEXTIDS-----	EKDLDDSDDD 1571	
Qy	521	EEAFARSAMIVHPMDLMDLAHANRSKPPYRRRSRSRSPER--	GOSPKRVNPEEK 577	
Db	1572	IEI-LEECIIISAMP-----TKSSHRAKKPAQOTSKLPPYARKPSOLPYUKLLPQNR	1623	
Qy	578	EREFRNFREVPYDIAKAY---GRSVFDPPEEKERYREMERKYRWYKKGVAVG	634	
Db	1624	LQP---QHVSTFPEDDDPRVYCEGPIPFSTATSLDIT-----	IESPNEILAAG 1672	
Qy	635	AQPPASNREDFSPERLLPLNTINSPTGREDYLAAGSHRNRLGNYPEKLISTDSH	694	
Db	1673	EGVAGGAQOGEFEKROTI P-----TGRSTDEAOG-----	GKTSSVTIPELDN 1716	
Qy	695	NAKN-----PKSKEKE-----	SENVPG---DCKGKKHKHR 723	
Db	1717	KAEBGDIILAECSINAMKGGSHKPFVKKIMDOYOGASASSAPKNOLDKRRKPTSPV	1776	
Qy	724	K-----RNEKEGESRFLNPELLETSRKGSSGIDETKDTLFLVPSRDA---	772	
Db	1777	KPIQNTETRYRRARKNADSKNNINAEIRVSDNKKDKONLAKNSOJFMDKLPNNEDRYG	1836	
Qy	773	-----TPVDEPM-----DAESITFKSVSDKDKREKDKPRVASDTKRRKSDGA	816	
Db	1837	SFAPDSPHHYPIEGTGYCFSRNDSSLSIDF-----	DDDDVDLSREKALERAKENKESEA 1892	
Qy	817	TAKDNVLYKSGKQEVVDODREKSPRSEPLKKAEEKITDSVK--	PSSSOKDER-- 872	
Db	1893	KVTSHTELTSNOOSANKTOAIAQINRQOPKPILOKOSTPQSSKDIJDGCAATDEKLO	1952	
Qy	873	---YTGPF-----	RAHNSAKDTR--ROSOPTRRSKRTVPTS 907	
Db	1953	NFAINTPVCNSHNSLSLSIDDOENKNEPIKETPPSQGEPSKPAQSGIARPSF	2012	
Qy	908	SQKSQPVTRRPRSLRKINY-----LI-----	AREKNEREK----- 938	
Db	2013	HVEDPVCFSNSSLSSLSIDSEDLLOECISSAMPKKKPRLKGDNKHSRPMGIL	2072	
Qy	939	-----	RKSYDK 945	
Db	2073	GEDLTLDKIDQRPDSEHGLSPSENEFMKAIOEGANSIVSSLHOAAAACLSROASSD	2132	
Qy	946	DF-----	ESSSMKISKVEGEIYKPSPKRKMGGVDEKLEPRPEKRI 987	
Db	2133	DSILSLKGISLSPFHILTPDOEKEPPTSXKGRILLKPEEKTLF--	TKIIPSEKGIK- 2189	
Qy	988	ASSTPAKKIKLNEETKKGNAENASTTEPEBEKLESTSSKIKOEKYGAKRRKVAAGE	1047	
Db	2190	-----GKKVYKSLITLTKVNSNBSISQOMQP---LQAMMPIS---	RGRMIHITPVR 2237	
Qy	1048	GSSSTLVLYTSTSTSGSPVR---KSEKTDTKRVIVKTYMEBYNDNTAPAEVYIIM-	1102	
Db	2238	NSSS-----TSPVSKKBPPLTKPASKSPSEGQATATSPGAKPVSXELSAPVARTSOIG	2293	
Qy	1103	---QVPQSKMKDKDFESEBEDEVKTQIQGVKPS-----	SIIMN 1139	
Db	2294	GSSKAPSRHSGRSDTSPRPAQOPLSRPIQSPGNSSISPERNGISPPNKILSQLPRTSPST	2353	
Qy	1140	VTTKRSARAK---YE-----	KESQOPKLOKLPREASHELMOHELSSKGSASS 1186	
Db	2354	ASTKSSGGKMSYISPGRHSQOONLTKQTGLSKNASSIPRSESASKGLGNNGGAKMK	2413	
Qy	1187	-EKGRAPKREHSSEKDNPK---RKSGAQDPKESTVDR--	LSQEGHKTLSQSG-- 1236	
Db	2414	VELSMTSSTKSSGSESDRSEKRPVLYVROSTIKRABPPTLRKLEBESASPELSISPSRRAS	2473	
Qy	1237	ETRTSEK-----	HESVGSSKNDFTPGDKRKVDYDSRDYSSSKRDERGE 1281	

[illegible]

```

US-08-821-355A-7
; Sequence 7, Application US/08821355A
; Patent No. 5851775
;
; GENERAL INFORMATION:
; APPLICANT: Barker, Nick
; APPLICANT: Clevers, Hans
; APPLICANT: Korinek, Vladimir
; APPLICANT: Morin, Patrice
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; APPLICANT: Sparks, Andrew
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC
; TITLE OF INVENTION: Interact to Prevent Cancer
; NUMBER OF SEQUENCES: 11
;
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20001
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,355A
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 514
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
;
; FILING DATE:
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A
; REGISTRATION NUMBER: 32,145
; REFERENCE/DOCKET NUMBER: 1107.05064
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: 97430 BMB UT
;
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2973 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; MOLECULE TYPE: NO. 5851775e
;
US-08-821-355A-7

```

```

Query Match Similarity 3.7%: Score 274; DB 2; Length 2873;
Best Local Similarity 18.5%: Pred No. 1.6e-09;
Matches 282; Conservative 188; Mismatches 612; Indels 446; Gaps 61;

Qy 173 FRYTVPCTCDKRTKAGSCDSCGTLRLPAPISISLTSNO--SSLAPVYSGNSSAPAPVP 230
      | : : : : | : : : : | : : : : | : : : : |
Db 1195 FSRKSSSSQSSSTTEHSSSESTNTS---TPSSNAKRONLHPSSAOSNRGGQOKA----- 1246
      | : : : : | : : : : | : : : : | : : : : |
Qy 231 DITATVTSIVHSERS-----DGPE--RDSDNKLLPAAA-----LTSERKSGASSTA 274
      | : : : : | : : : : | : : : : | : : : : |

```

```

Db 1247 ---ATCKSSINQETIQYCVEDTPICFSCSSLSLSSADELICNQTOEADANLTLQ 1303
Qy 275 ITALMEERG-----VPGTSP-----WNSTFVQSLHGLIPIITG---FVR 312
Db 1304 IAEIKETIGTRSAEDPVSEVAVSQHPRTKSSRLQSSLSSESAHKKAVERSQKAKSSDK 1363
Qy 313 INARPGGGRGWESHNKLGLVSPPOQIRRGERSCYRSINNGRHHSRROQTOSPSPA 372
Db 1364 SGAQTP-----KSPFHYVQETPLMFESRCHTSVSLDSFESRSIASSVQS 1407
Qy 373 TPCEVPV-----PPPLYPHPHTLPLPPGVPPQFSP---QFSPSQPTAGYVSP 420
Db 1408 EPCSGMVGIISPSDLPSQPTMPSPRSKTPP---PPQATQTKREYVKNNAPTAKRES 1465
Qy 421 PPG-----FPRAA---NISTACFSPGYTASHNTMPTQAPLSREFFYR--- 463
Db 1466 GPKQAAVAAVOYOVLEDAQTLHLHFAESTPDPGFCSSLSLALSLDPEFLOKDYELRIM 1525
Qy 464 ---BONDGRSKPPYSGSYSRSSYTDSOGLAQHIALTLSPSAHTLDDLHHPHP 520
Db 1526 PPVOENDGNETE-----SEQPKESNEQEKAEKTIIDS-----EKDLLDSDDD 1571
Qy 521 EEAARSAMIHMDLMDIAHARSPPYRRYRSRSRSPPEFR---GOSPTKRNVPREK 577
Db 1572 IEL-LEECIIISAMP-----TKSSRKAKKPAQTASKLPPYARKPQSLPYKLLPSQNR 1623
Qy 578 EREFENRREVPYPPYDIAYV---GRSVDFNDPEFEKERYREMYEKRYGYAVG 634
Db 1624 LQP---QKHVSFTPGDDMPRYVCVECTPIINFSTATSLDLT-----IEPPNELAAG 1672
Qy 635 AOPRSARREDFSPERLLPLINRNSPFRGRREDYAAGSHNRMLGNGNYPEKILSTRDSH 694
Db 1673 EGVNGAGSGEFEKRTIP-----TEGSTDEAQ-----GKTSSVTIPELDN 1716
Qy 695 NAKDN-----PKSKEKE-----SENVPG---DGKNKHKHR 723
Db 1717 KAEEDILAEICINSAMPKSGHKKPRVKKINDQVOQASASSAPRNQLDGKKKPTSPV 1776
Qy 724 K-----RNEEKGESSEFLNPELLETSKRCGSSGIDETKTDTLFLVLPBRDA--- 772
Db 1777 KPDPQNTERTVRKKNADSKNNLNAERFVSNDKSKKONLKNNSKDFNKLPPNEDRYVG 1836
Qy 773 -----PPVREPM-----DAESITFKSVSDKDKREKDKPKVKSDDTKRKSDGSA 816
Db 1837 SPADSPHHYPIEGTPICFSRNDSLSLDF---DDDDVDLSREKAELRKAKENKESEA 1892
Qy 817 TAKDNVLKPSKGQEKYDGDREKSPRSEPLKKAKEBATKIDSVK---PSSSSQKDEK-- 872
Db 1893 KVTSHTELTSNQOSANKQATAKQINRGOPKPILOKOSTPOSSKDIPIRGAATDEKLQ 1952
Qy 873 ---VTGTP-----RKANSKSAKTR---ROSQPTRRSKRTVPKTS 907
Db 1953 NFAIENTPVCFSHNSLSLSDIDQENNNKENPIKETEPDPOGEPSKQASGYAPKSF 2012
Qy 908 SOKSOPVTRRRPSRLKINYN---LI-----AREKNEREK----- 938
Db 2013 HVEDTPVCFSKNSLSLSDISEDDLQECISSAMPKKKKPSRLKGDNKHSPPRMGGTL 2072
Qy 939 -----RKKSYDK 945
Db 2073 GEDLTLDKDIORPDEHGLSPDSENFDMKAIQEGANSIVSSLHQAAAAACLSRQASSDS 2132
Qy 946 DF-----ESSMKIKVEGTEIVKPSPKRMKGDVKELERTPEKDKI 987
Db 2133 DSIILSKGSIIGSPFHLTPQOEKPFITSNKGPRILKGEKSTLE--TKTIESESGIK- 2189
Qy 988 ASSTPAKKIKILNRETGKIGNAENASTKEPSEKLESTSSKIQEKVKGAKRKRVAGSE 1047
Db 2190 -----GKKVKYKSLITGKVRNSSEISGOMKOP---LQANMPSIS---RGRTHIHPGVR 2237
Qy 1048 GSSSTLVDTYSTSTSGSPVR---KSEKTDTKRTVIKTMEYNNNDNTAPAEDVIIMI- 1102

```

```

Db 2238 NSSSS-----TSPVSKKGPPLKTPASKSPSEGOATATTPRGAKPSVKSSELSPVARTSOIG 2293
Qy 1103 ---QVPQSKMMDKDFEESPEEDVKTTOPIQSVGKPS-----SIITKN 1139
Db 2294 GSSKAPSRSGSKDSTPSRPAQOPLSRPIQSPGRNHSISPERNGISPPNKLSQLPTSPST 2353
Qy 1140 VTTKPSATAK--YTP-----KESQEPKLOKLPKKEASHLMOHELRSKGSASS 1186
Db 2354 ASTKSSGSGKMSYTPGQMSQONLTQOTGLSKNASSIIPRSRSASKGJLNMNNGAKKK 2413
Qy 1187 -EKGRADREHSGSEKDNPKD---RKSGADPKESTYDR--LSQGHFKTILSQSK--- 1236
Db 2414 VELSRMSTKSSGSEBSRSEPVILRQSTFIKEADSPILRKLREBSASFESLSPSRPAS 2473
Qy 1237 ERTISEK-----HESVRGSSNKDFTPGDKDKYDSDRYSSSKRDERGE 1281
Db 2474 PTRSQAQTPVLSPLPDMSLSTHSSVQAGWKKLPPNLSPTLEYN--DGRPAKRHD--- 2527
Qy 1282 LAR-RKDSPPRGKESLSG--OKSKLREERDLPKKAESKKSNSPPRDKKPHDHKAPYET 1338
Db 2528 IARSHSESPSLPINRSGTWKREHSHSSSLPRVSTWRTGTSSSILNASSPSE----- 2582
Qy 1339 KRPCEETKPPVDKNSGKEREKHAAREARNG 1366
Db 2583 KAKSEDEKHVNSISGTKQSKENQVSAKG 2610

```

Search completed: September 12, 2002, 19:10:23
 Job time: 11187 sec

...

OM of: US-09-811-045a-1 to: GenBank: * out_format : pfs
Date: Sep 12, 2002 5:11 PM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framer-p2n.model -DEV=x1h
-O=/c9n2_1/USPTO.spool/US09811045/runtac_12092002_132242_5982/app-query.fasta_1.1485
-DB=GenBank -FORMAT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -EGAPOP=6.000
-DEEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEADSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09811045_@cgn1_1.5886
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPRXY -WAIT -THREADS=1

Search information block:

Query: US-09-811-045a-1

Database: GenBank: *

Database sequences: 1797656

Search length: 1873333701

Search time (sec): 3967.930000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
gb_ro:MM083913	+ 6724.00	2941.98	1.7e-155	5131	! U83913 Mus musculus proliferation
gb_ro:MM028789	+ 6638.00	2904.66	2.0e-153	5191	! U28789 Mus musculus p53-associ
gb_pr:AC352051	- 5739.00	2512.62	1.4e-131	5376	! AF352051 Homo sapiens prolif
gb_pr:AC010321	- 4135.00	1788.77	4.0e-88	3011	! X85133 H. sapiens RBQ-1 mRNA
gb_pr:HSR801	- 3938.00	1732.21	4.0e-88	3011	! X85133 H. sapiens RBQ-1 mRNA
gb_ro:AC08487	- 2665.00	1149.41	1.2e-55	175948	! AC08487 Mus musculus clone
gb_hcg:AC099701	+ 2576.00	1109.22	2.0e-53	222349	! AC099701 Mus musculus clone
gb_pat:AK048215	+ 1838.00	821.32	2.2e-37	1800	! AR048215 Sequence 11 from pat
gb_hcg:AC095130	+ 1777.00	763.58	3.9e-34	23479	! AC095130 Rattus norvegicus c
gb_pr:AK026954	- 1750.00	781.01	3.9e-35	2320	! AK026954 Homo sapiens CDNA: FI
gb_hcg:AL670951	- 1585.00	678.11	2.1e-29	240937	! AL670951 Mus musculus chromo
gb_pr:HSR802662	+ 1353.00	609.25	1.4e-25	2026	! AL359564 Homo sapiens mRNA: CI
gb_pr:BO15318	+ 1046.00	482.33	1.7e-18	740	! BO15318 Homo sapiens, clone 1
gb_hcg:AC095130	- 907.00	384.64	4.6e-13	153479	! AC095130 Rattus norvegicus c
gb_hcg:AL450323	- 773.00	328.59	6.1e-10	110015	! AL450323 Mus musculus chromo
gb_hcg:AC105975	+ 676.00	289.87	8.8e-08	68409	! AC105975 Mus musculus chromo
gb_hcg:AC04642	- 544.00	252.04	1.1e-05	3937	! AF132177 Drosophila melanogast
gb_hcg:AC020509	- 530.00	220.88	0.0006	148432	! AC020509 Drosophila melanogast
gb_hcg:AC012167	- 530.00	220.17	0.0007	157851	! AC012167 Drosophila melanogast
gb_hcg:AC003463	+ 530.00	215.98	0.0011	164443	! AC003463 Drosophila melanogast
gb_hcg:AC105428	- 500.00	208.85	0.0029	127839	! AC105428 Mus musculus clone
gb_hcg:CE36F2	- 432.00	191.04	0.0281	22747	! Z81532 Caenorhabditis elegans
gb_ro:AF138789	+ 417.00	187.49	0.0443	14825	! AF138789 Rattus norvegicus mu
gb_ro:AF227534	+ 417.00	186.97	0.0473	15961	! AF227534 Rattus norvegicus mu
gb_pat:AK06459	+ 413.00	198.18	0.0112	2365	! AK06459 Sequence 4 from Paten
gb_hcg:AC017682	+ 412.00	174.21	0.2431	75313	! AC017682 Drosophila melanogast
gb_hcg:AC079916	+ 410.00	168.45	0.0068	206258	! AC079916 Homo sapiens chromo
gb_hcg:AC063368	+ 408.00	168.45	0.0068	129889	! AC063368 Homo sapiens chromo
gb_pr:HMUMLAF	+ 405.00	187.76	0.0427	6940	! L22179 Human MTL-AF4 der(11) t
gb_pr:AP002057	+ 405.00	171.12	0.3612	73391	! AP002057 Arabidopsis thaliana
gb_hcg:AC003420	+ 401.00	159.51	1.60	302835	! AC003420 Drosophila melanogast
gb_pat:AX210662	+ 400.00	183.75	0.0715	9027	! AX210662 Sequence 304 from Pat
gb_pr:AB016092	+ 398.00	179.02	0.0715	9027	! AB016092 Homo sapiens mRNA for
gb_ov:GG119187	+ 398.00	179.02	0.0715	15717	! Y19187 Gallus gallus mRNA for
gb_hcg:AC084463	+ 396.00	166.13	0.6855	85644	! AC084463 Caenorhabditis briggs
gb_hcg:AC004441	+ 395.00	160.47	1.42	181356	! AC004441 Drosophila melanogast
gb_pr:AF201422	+ 394.00	181.99	0.0901	7789	! AF201422 Homo sapiens splicing
gb_pr:AC093517	+ 394.00	160.88	1.34	160601	! AC093517 Homo sapiens chromo
gb_hcg:AC092117	- 394.00	159.06	1.70	208529	! AC092117 Homo sapiens chromo

gb_in:DMC49E4 + 393.00 170.17 0.4084 39749 ! AL031128 Drosophila melano
gb_hcg:AC026106 - 392.50 160.06 1.49 164485 ! AC026106 Homo sapiens chr
gb_hcg:AC097329 - 390.00 157.62 2.04 199706 ! AC097329 Pan troglodytes
gb_pr:AC004493 - 389.00 167.77 0.5553 43661 ! AC004493 Homo sapiens chro
gb_in:CBRC45011 - 388.00 170.76 0.3785 26703 ! AC084652 Caenorhabditis br

Seq_name: gb_ro:MM083913

seq_documentation_block:

LOCUS MM083913 5131 bp mRNA linear ROD 10-NOV-1998

DEFINITION Mus musculus proliferation potential-related protein (P2P-R) mRNA,

complete cds.

ACCESSION U83913

VERSION U83913.1 GI:3858884

KEYWORDS

SOURCE

ORGANISM

house mouse.
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 5131)

Witte,M.M. and Scott,R.E.
The proliferation potential protein-related (P2P-R) gene with

domains encoding heterogeneous nuclear ribonucleoprotein

association and Rb1 binding shows repressed expression during

terminal differentiation

Proc. Natl. Acad. Sci. U.S.A. 94 (4), 1212-1217 (1997)

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

FEATURES

source

gene

CDS

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

QSKWKKDFESEEEDVKTTPIQSVGKPSIIINVTTPKPSATAKYTEKESEDEKLOK
 LPEKASHELMOHIELRSKSGASSEKGRKADREHSGKDPDKSGAQPKSKSTYDR
 LSEOGHETKLISOHKEPTSEKHESEVSGSNKEFTPRDKVDSYDSYSSKRADER
 GELARRKDSPPRKESLSGOKSLREERDLPKGAESKSNSSPRDKKPHDKAPYE
 TKRPECERTPVDKISGEREKHAAERANGKDSGKRLPCLINPDLMEKELAVCOYE
 KSAVKRPQLSHSSRLSSDLTRETDEAEPDYNESSESNSVKEEAVASISKDLK
 EKTETAKRESULTVATASQPGADRSQSSPSVPSRSHSQSTRSHSSASASQSD
 DSKKKKKKKKKKKKKKKKKKKKKHAGADGDVEKSOQHKKKKKKKKKKKKKKKKKKKK
 OKVRSYTVV"

BASE COUNT 1728 a 1165 c 1192 g 1046 t
 ORIGIN

alignment_scores:
 Quality: 6724.00 Length: 1409
 Ratio: 5.010 Gaps: 6
 Percent Similarity: 95.245 Percent Identity: 93.329

alignment_block:

US-09-811-045a-1 x MM083913 ..

Align seg 1/1 to: MM083913 from: 1 to: 5131

1 MetMetGluValLysAspProAsnMetLysGlyAlaMetLeuThrAsnTh 17
 |||||
 139 ATGATGGAAGTGAAGATCCCTTAACATGAAGAGTGCAATGCTTACAAACAC 188
 |||||
 17 rGlyLysTyrAlaIleProThrIleAspAlaLysAlaTyrAlaIleGlyL 34
 |||||
 189 TGGAAATATGCAATACCAACTATATAGATGACAGAGCTTATGCAATCGGGA 238
 |||||
 34 yLysGluLysProProPheLeuProGluGluProSerSerSerSerGlu 50
 |||||
 229 AGAAGAGAAACACACCTTCTTACAGAGAGCCATCATCATCTTCACAGA 288
 |||||
 51 GluAspAspProIleProAlaGluLeuLysCysLeuIleCysLysAspI 67
 |||||
 289 GAAAGATGATCCCTATCCAGACAGAGCTTGTGCTCATCTGCACAAAGACAT 338
 |||||
 67 eMetThrAspAlaValIleIleProCysCysGlyAsnSerSerCysAspG 84
 |||||
 339 CATGACTGATGCTGTGTCATTCCTGCTGCTGGAACAGTACTGATGAG 388
 |||||
 84 LuCysIleAaTThrThrLeuLeuGluSerAspLysHisThrCysProThr 100
 |||||
 369 AATGATATAAGAACACACACTCTTGAGTGCAGATGAACATATCATGTCCACA 438
 |||||
 101 CysHisGlnAsnAspValSerProAspAlaLeuIleAlaAsnLysPheLe 117
 |||||
 439 TGTCAACCAAAATGATGTTCTCTCATGCTTAAATGGCCAACACAGTTT 488
 |||||
 117 uArgGlnAlaValAsnAsnPhelLysAsnGluThrGlyTyrThrLysAspL 134
 |||||
 489 AGGACGGGCTGTTAATACTTAAATAAGAACTGCTATACAAAACGAC 538
 |||||
 134 euArgLysGlnLeuProProPheLeuPheLeuValProProProArgPro 150
 |||||
 539 TACGAAAACAGTTACCTCCACCCACCCACCCAGTACACACCAAGACCA 588
 |||||
 151 LeuSerGlnAAsnLeuGlnProArgSerArgSerProIleLeuArgI 167
 |||||
 589 CTCAGCAGCGGACCTGACGCTCTGATGAGATCTCCCATATCAAGACA 638
 |||||
 167 nGlnAspProValPheArgTyrThrValSerProThrCysSerAspT 184
 |||||
 639 GCAGGATCTCTGATGATTCACAGTACATCGTCTCCGCTCAGTCA... 684
 |||||
 184 hrLysThrAlaGlySerCysSerAspSerGlyThrLeuSerArgLeuPro 200

 684
 201 AlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAlaProPr 217
 |||||

665 GCTCCCTCTAATATCTTCATTAACCTTCAATCCATCTGCCCTGGCTCTTC 734
 217 oValSerGlyAsnProSerSerAlaProAlaProValProAspIleThrA 234
 |||||
 735 TGTGCTGTGAAATTCATCTCTTGCCCCAGCTCCATGATGATATACAG 784
 |||||
 234 laThrValSerIleSerValHisSerGlyLysSerAspGlyProPheArg 250
 |||||
 745 CAACCGTGTATTATACAGTCCACCTCAGAAAATTCGATGAGACCTTTCCG 834
 |||||
 251 AspSerAspAsnLysLeuLeuProAlaAlaAlaLeuThrSerGluHisSe 267
 |||||
 835 GATTCGTGAATTAATTAATTTATGCCACCTGCGCCCTTACATCGAACAATTC 884
 |||||
 267 rLysGlyAlaSerSerIleAlaIleThrAlaLeuMetGluGlyLysGlyV 284
 |||||
 885 AAAGGAGCCTCTTCAATTTGCTATTACCTCTTATGGAAGAAAAGGGGT 933
 |||||
 284 aLProGlyThrSerProThrAsnSerIlePheValGlyGlnSerLeuLe 300
 |||||
 934 A.CCAGGTACACAGTCCCTTGGAACCTCCATCTTGTGGACAGTATTAT 983
 |||||
 300 uHisGlyGlnLeuIleProThrThrGlyProValArgIleAsnAlaAla 317
 |||||
 984 ACATGAGACAGTTGATTCACACACTGGCCAGTAAAGAAATCAATGCTGCTC 1033
 |||||
 317 rGProGlyGlyValrProGlyTyrGluHisSerAsnLysLeuGlyTyr 333
 |||||
 1034 GTCCAGGTGTGTGGCCGCGCAGCTGGAGACATTTCCACACAGTTTGGGTC 1083
 |||||
 334 LeuValSerProProGlnIleArgArgGlyGluArgSerCysTyrAr 350
 |||||
 1084 CTAGTTTCTCCACACAGCAAAATTAAGAGAGAGAAAGACGTGTACAG 1133
 |||||
 350 gSerIleAsnArgGlyArgHisHisSerGluArgSerGlnArgThrGln 367
 |||||
 1134 AAGTATTAACCGCGGCGCAGCACACAGCGAAGATCACAGAGACTCAAG 1183
 |||||
 367 erProSerLeuProAlaThrProCysPheValProValProProPro 383
 |||||
 1184 GCCCATCACTTCCACACMACTCCAGTCTTTGTCCGCTTCCACACTCTCT 1233
 |||||
 384 LeuTyrProProProProHisThrLeuProLeuProProGlyValProPr 400
 |||||
 1234 TTGTATCCGCGCTCTCCCATATCACTCTCTTCTCCAGGTGATCTCC 1283
 |||||
 400 OProGlnPheSerProGlnPheProSerSerGlnProProThrAlaGlyT 417
 |||||
 1284 TCCACAGTTTCTCCTCAGTTTCTCTGGCCAGCCTCCAAACAGCAGAT 1333
 |||||
 417 yrSerValProProProGlyPheProProAlaProAlaAsnIleSerThr 433
 |||||
 1334 ATAGTCTCCTCTCCAGAGATTTCACACAGCTCTGCCAATATATACACA 1383
 |||||
 434 AlaCysPheSerProGlyValProThrAlaHisSerAsnThrMetProTh 450
 |||||
 1384 GCTTCTCTTTCACAGGTGTCTCCACTGCCCATTAATAATACATGCCAC 1433
 |||||
 450 rThrGlnAlaProLeuLeuSerArgGluGluPheTyrArgGluGlnAsna 467
 |||||
 1434 AACACAAACACCTCTTTGTCCAGGGAAAGAAATTCATAGAGAGCAAAAAC 1483
 |||||
 467 sPlyGlyLysGluSerLysPheProTyrSerGlySerSerTyrSerArg 483
 |||||
 1484 GACTAAAGGAAGAGTCTAATTTCCCTATAGGGGTCTCATGATTCACAAGA 1533
 |||||
 1534 AGTCATACACTGACTCAAGTCAAGGTCTGGCTCAACACAGTTCCAGCTC 1583
 |||||
 500 euThrLeuSerProSerAlaAlaHisThrLeuAspLeuLeuHisAspHis 516
 |||||
 1584 TTACTCTGCTCTTTCACGCGCTCAGATTCTGCTCTATTTCACGATCAC 1633
 |||||

517	ProHisProIoGluGluLualGluLaaArgSerAlaMetIleValHisMe	533
1634	CCCCATACCCCGAGGAGGCGAGGCGCAAGACCGCAATTCAGCTTGACGG	16833
533	troAspSerMetLarPLeaIleHisAlaArgSerArgSerProProtyrA	550
1684	TCAGATGTCACGCGATACCCACCGATCTTAAGGTCAAGGCTCCCTCATTA	17322
550	rgArgTyraArgSerArgSerArgSerProProGluLupPhaArgGluGlnSer	566
1733	GACGATACACACTCAGCGGTCCAGATGTCCTCAAGCATTTAGGGGACAGCT	17822
567	ProThrLysAlaGlnValProArgGluGluLysGluLysGluArgGluTyrrPheAs	583
1783	CCCACTAAACGCAAAAGTCACTCGAGGAAACACAGCGCTGAGTATTTTAA	18322
583	narGtyrArgGluValProProProTyrrAspIleLysAlaTyrrGlyA	600
1833	TGATATACGAGAAAGTCCACCCCTTAAGACATCAAAACCTATTATGTGGC	18822
600	rgSerValAspPheArgAspProPheGluLysGluLysGluArgTyrrArgGluTrrP	616
1883	GGAGGTGCACTTTAGAGACCCATTTGGAGAAAGAACGCTACCGGGAAAGG	19322
617	GluArgLysTyrrArgGluTrrPtyGluLysTyrrTyrrLysGlyTyrrAlaVala	633
1933	GAAAGGAATACCGGAGAGGTGATGAGAAAGTACTCAAAAGGATACGGCGGT	19822
633	IcLysAlaGlnProArgProSerAlaAsnArgLysAspPheSerProGluA	650
1983	GGAGGCTCAACCTAAACCTCGCCCAATGAGAGGACCTTTCTCCAGAGA	20322
650	rgLeuLeuProLeuAsnIleArgAsnSerProPheThrArgGlyArgArg	666
2033	GACCTTGACTTATATACAAATTCACCTTTACAGAGGCGCGCAGA	20822
667	GluAspTyrrAlaAlaGlyLysSerHisArgAsnArgAsnLeuGlyGlyAs	683
2083	GAAAGCTATAGCTGCGAGACAAAGTCATGAAATATAGAAATTCAGGTGCAG	21322
683	nHyrrProGluLysLeuSerThrArgAspSerHisAsnAlaLysAspAsn	700
2133	CTATCCAGAAAGCTTTCAACAAGGACACAGTCACATCCAAAGATTAATC	21822
700	roLysSerLysGluLysGluSerGluAsnValProGluLysGlyLysGly	716
2183	CAAAATCGAAGAGAGAGAGAGAGTGCAGATGTTCCAGGAGACGCGCAAGGG	22322
717	AsnLysHisLysAlaHisArgLysArgArgAsnGluGluLysGlyGluGlu	733
2233	AACAAGCATTAAGAAACACAGAAACGAAAGAAACGAAAGGCGGAGAGA	22822
733	userGluSerPheLeuAsnProGluLeuLeuGluLysSerArgLysCysA	750
2283	GAGTGAGACCTTCTCGAACCAGAGTACTGCGGAAGTTTAGGAAATGCA	23322
750	rgLysSerSerLysLysLeuAspGluTrrLysThrAspThrLeuPheValLeu	766
2333	GGGGAATCGTCACGAGATGTGTAAAGAAAGACAGATACCTGTGTGTTTTTC	23822
767	ProSerArgAspAspAlaThrProValArgAspGluProMetAspAlaGlu	783
2383	CCAGAGAGAGAGATGCAACACCTTTAGGGATGAGCCAAATGGAGCGCAGA	24322
783	uSerIleThrPheLysSerValSerAspLysAspLysArgGluLysAspL	800
2433	ATCGATCACTTTCAAGTCAGTATCTGACAAAGACAAAGGCGAAAGGATA	24822
800	ysProLysAlaLysSerAspLysTrrLysAlaGlyLysSerAspGlySerAla	816
2483	ACCCAAAATATAAAAGTGCACAGACCAACGCAAAAGGAGCGGGCTCTCT	25322

817	ThrAlaLysLysAspAsnValLeuLysProSerLysGlyProGlnGly	833
2533	ACAGCCAGAAAGAACAAATGTTTTAAACCTTCAGAAAGCACTCAAGAAA	2582
833	SValAspGlyAspArgGlyLysSerProArgSerGlnProLeuLysL	850
2583	GGAGAGTGGAGACCGGAAAAATCTCTCGGCTGAGCGGCACTCAAA	2632
850	ysAlaLysGlnLbAlaThrLysIleAspSerValLysProSerSer	866
2633	AAGCCAAAGAGAGGCTACAAAGATTTGACTGTGAAAACCTTCCTCGCT	2682
867	SerGlnLysAspGlyLysValThrGlyThrProArgLysAlaHisSerLys	883
2683	TCACAAAGAGTGAAGAGAGTCTACTGGAACCCCTAGAAAAGCCATTCTAA	2732
883	sSerAlaLysAspArgThrArgArgGlnSerGlnProArgThrArgArgSer	899
2733	ATCTGCAAAAGAACACCCAGAGGAGCAAGCCAGCAAGAGCAAGAGTCA	2782
900	LysArgThrValProLysThrSerSerGlnLysSerGlnProValArgThr	916
2783	AAAGAGACTGTTCCAAAGACATCAAGTACAGAAAACCGCCAGTAAGAC	2832
916	ArgArgProArgSerLeuArgLysIleAsnTyrLeuIleAlaArgGlu	933
2833	GAGAAAGCCCAAGACCGTGGAGAAAATAACTACTTGTATGACAGGGAGA	2882
933	ysAsnGlnArgGlyLysArgGlyLysSerAlaAspLysAspPheGly	949
2883	AAAGCAAAAGAGAAAACCGAAGAAAAGAGTAAATAAAGTTTTGAGT	2932
949	erSerSerMetLysIleSerLysValGlnGlyThrGlnIleValLysPro	965
2933	CGTCTCATATGAATCTCTAAAGTAGAAGAGAACAGAAATAGTAAACA	2982
966	SerProLysArgLysMetGlnLysAspValGlyLysLeuGluArgThrPr	982
2983	TCACCAAAACGGAAAATGGAAAGTGATGTGAAAAGCTGGAAGAGACCC	3032
982	oGlnLysAspLysIleAlaSerSerThrThrProAlaLysLysIleLysL	999
3033	AGAAAAGGCAAGATTGCATCTCACTACTCCACCCCAAAAATCAAAAC	3082
999	euaAsnArgGlnThrGlyLysLysIleGlnAsnAlaGlnAsnAlaSerThr	1015
3083	TCACACGAGAAACTGGAAAAAAATTTGGAAATGCAGAAAATGCACTACT	3132
1016	ThrLysGlnProSerGlnLysLeuGlnSerThrSerSerLysIleLysGln	1032
3133	ACAAAAGAACCTCTGAAAAAATTTGGAGTCAACATCTACCAAAATCAACA	3182
1032	nGlnLysValLysGlyLysAlaLysArgLysValAlaGlySerGlnLys	1049
3183	GGAAAAAGTCAGAGGAAAGCCAAACGGAAATACCTGGGTCGGAAGCT	3232
1049	erSerSerThrLeuValAspTyrThrSerThrSerSerThrGlyLysSer	1065
3233	CCAGCTCCACGCTTGAGGATTCACACAGACAGTCAAGTCACTCGAGGACGT	3282
1066	ProValArgLysSerGlnGlyLysThrAspThrLysArgThrValIleLys	1082
3283	CCGTGTAGGAATCTGGAAGAAAAAGACAGATTCACAGCGAACGTCATTAA	3332
1082	sThrMetGlnGlyTyrAsnAsnAspAsnThrAlaProAlaGlnAspValI	1099
3333	AACTAAGGAGGATATTAATAAAGATTAACACAGCTCTCTGAGAGTGTTA	3382
1099	IleIleMetIleGlnValProGlnSerLysTyrAspLysAspAspPheGlu	1115
3383	TAAATTAATGATCCACGTTCTCTGCTCAATGGAATGAAGATGACTTGTAG	3432
1116	SerGlnGlnGlnAspValLysThrThrGlnProIleGlnSerValGlyLys	1132

|||||
3433 TCTGAGAGAGATGTTAAACCAACACCTTACAGACTGAGGAA 3482
1132 sProSerSerIleIleLysAsnValThrThrLysProSerAlaThrAla 1149
3483 ACCATCGAGTATATATAAAATGTCTACTACTAACCCATCGGCTACGGCTA 3532
1149 YSTYrThrGluLysGluSerGluInProGluLysLysLysLysLysLysLys 1165
3533 AGTACACCGAG 3582
1166 LysGluLysSerHisGluLysMetGlnHisGluLysLysSerLysGlu 1182
3583 AAGAGAGGAG 3632
1182 YSerAlaSerSerGluLysGluLysAlaLysAspArgLysHisSerGlys 1199
3633 CAGTGCCTCCAGTGAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3682
1199 eRgIuLysAspAsnProAspLysArgLysSerGlyAlaGlnProAspLys 1215
3683 CGGAG 3732
1216 GluSerThrValAspArgLysSerGluGlnGlnLysHisPheLysThrLys 1232
3733 GAGAGACAGTGTGACCGCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 3782
1232 rGlnSerSerLysGluThrArgThrSerGluLysHisGluSerValArg 1249
3783 TCAGTCTCTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3832
1249 YSerSerAsnLysAspPheThrProGlyArgAspLysLysValAspTyr 1265
3833 GTTCTCTCAATAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3882
1265 AspSerArgAspTyrSerSerSerLysArgArgAspGluArgGluLys 1282
3883 GACAG 3932
1282 uAlaArgArgLysAspSerProProArgGlyLysGluSerLysSerGly 1299
3933 AGCAG 3982
1299 LysSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1315
3983 AGAAG 4032
1316 SerLysLysSerAsnSerSerProProArgAspLysLysProHisAsp 1332
4033 TCTTAAAAAATAATCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4082
1332 sLysAlaProIleGluThrLysArgProGlyGluLysLysLysLysLysLys 1349
4083 TAAAG 4132
1349 sPlysAsnSerGlyLysGluLysLysLysLysLysLysLysLysLysLys 1365
4133 ATAAATATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4182
1366 GlysGluLysSerGlyLysLysLysLysLysLysLysLysLysLysLys 1382
4183 GGGAG 4232
1382 hLysProTyrArgArgSerThrLysLysLysLysLysLysLysLysLysLys 1398
4233 CCTTCCCATGAG 4282
1399 SerSerArgAsnProSer 1404
4283 TCAAGCGAG 4300
seq_name: gb_ro:MMU28789

seq_documentation_block:
LOCUS MMU28789 5191 bp mRNA linear ROD 14-MAR-1997
DEFINITION Mus musculus p53-associated cellular protein PACT mRNA, partial
ACCESSION U28789
VERSION U28789.1 GI:1546778
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE
AUTHORS 1 (bases 1 to 5191)
1 Simons, A., Melamed-Bessudo, C., Wolkowicz, R., Sperling, J.,
Sperling, R., Eisenbach, L., and Rotter, V.
PACT: cloning and characterization of a cellular p53 binding
protein that interacts with Rb
Oncogene 14 (2), 145-155 (1997)
TITLE
JOURNAL
MEDLINE 97163396
REFERENCE 2 (bases 1 to 5191)
AUTHORS Simons, A.H., Eisenbach, L., and Rotter, V.
Direct Submission
JOURNAL Submitted (08-JUN-1995) Arnold H. Simons, Cell Biology, Weizmann
Institute of Science, Rehovot, 76100, Israel
FEATURES
source
CDS
/db_xref="GI:1546779"
/translation="KDPNMKGAMLNTWTKYAIPITDAEVAATGKKEKPFLEPERSSS
SEEDPFPPELLCLICQIDMTDAVAVYPCGNSYDCICITALEDEHPCPCFCHONDV
SPDAILANKELRGNVNFENKNGYTRKRLKOLPPPPPPPPPPPPPPPPPPPPPPPP
SHQODPLMPTVYSSSHASHPASSISLTSNLSALAPSVSGPSSAPAPPPDITATVSISS
HSEKSDGPPRDSNDKLLPAALITSEHSKASSTATITLMEKGTQVPLVGLTSLGOS
LHAGQLIPTTGPVRINAARPGGGRGPEHSHNKLGYLVSPQOIRKGRSCYRSINGR
HHSERQORTGSLPATPVFVPPVPPPPPLPPLPPLPPLPPLPPLPPLPPLPPLPPL
AGYSVPDGFPPAPANIPTPVVSGVQWHSMTIPTQAPLSRSEFYRQRLKEE
KKSKLDEFTNDFAKELMEYKKIOKERRSFERSKSPYSGSSYSRSYTSKRSRGSST
RSRSYSRSFSRSYSRSYSPYPRGRGKSNYSRSYSHGYSRHSRSYSPYPRYRHS
RSRSYPOAFRGQSPTRKNVPQGETEREYFNRYREVPPIYDIKAYIGRSYDFRPFER
YREWERKYREWEYKYYKGYAQAOPPSANREDFSERLLPLNIRNSPFRREDYA
AQOSHNRNLGNGYPEKLESTRDSHNAKNDPKSEKSESEVNPQDGGNKKRKRKRK
ESESERLNRLNETSREKRESSGIDETKDTLFLVPSRDAVPVDEPMDASLTFKS
VADKDRKENDKRVKSDTKRKSDGATATKKNVLPKSPQEKYDGRDEKSPSP
LKKAEARAKIDSVKPSSSOKDEYVTPGPRKASHGSAKEHOEAPKADKEVKKCSK
DIKSEKPAKDKRAKPEKKNLLDSGKERKRTKEKSDKDPRESSMKISVVEGEI
VPSKPRKMEGVEKLETPERDKLASSTTPAKKIKLNRETGRKIGNAENASTTEPS
EKLESTSKIKOEKYGAKARKVAGSESSYLVDYTSSTGSGVRSSEKTDYKR
TYIKTWEENYNNNTAPAEDEVIIIMIQVQKMDKDESEDEBKVTQI0SYGKSS1
IKNVTTEKPGATATKYTESEOEPEKLOLKEKSHLMEHLEKSSKSSASSEKGRADR
HSHSGSEKNDKDRKSGAOPDKESTYDRLESQGHFKTLISQSKETRSSEKHEEDRSN
KDPTEGRDKKVVYDSDYSSSKRRDRGRLARKKSPSPGKESLISQKRSLEEDLP
KKGAEKSKSNSSPPRDKKPHDKAPETKRCPEKRPVYKNSGKEKEMAAARNGKE
SSGGKILPCIIPNPDPPEMERELAGOVEKAVPRQLSHSNLSLSDLTRETEAEP
DYNESDESNNVYKEEAAVASISKDLKERTERAKESLIVATASQDADRSOSQSPS
VPSRSHSPSGSQTSHSSASASGSDSKKKKKKKKKKKKKKKKKKKKKKKKKKK
DYEKSHKHKHKK
1540..1813
/note="encodes serine/arginine rich region"
BASE COUNT 1763 a 1183 c 1198 g 1047 t
ORIGIN
alignment_scores:
Quality: 6638.50 Length: 1439

Ratio: 4.976 Gaps: 9
Percent Similarity: 92.703 Percent Identity: 90.966

alignment_block:
US-09-811-045a-1 x MMU28789 ..

Align seg 1/1 to: MMU28789 from: 1 to: 5191

5 LysAspProAsnMetLysGlyAlaMetLeuThrAsnThrGlyLysTyrAl 21
|||||
2 AAGAGCTCAATATGAAAGGTGCAATGCTTACAAACACATGGAATAATATC 51
21 alleProThrIleAspAlaGluAlaTyrAlaIleGlyLysGlyLysP 38
|||||
52 AATACCAATATATGATGACAGAGCCATATGCATCGGAGAAAGAAAC 101
38 robProHeuLeuProGluLeuProSerSerSerSerGluLysAspPro 54
|||||
102 CACCCCTTCTACAGAGAGCATCATCTTCAGAGAAGATGATCCT 151
55 IleProAlaGluLeuLeuLysLeuIleCysLysAspIleMetThrAspAl 71
|||||
152 TTCCCGACAGAGCTCTGTGCTCATCTGCAAGACATCATGATGATG 201
71 aValValIleProCysCysGlyLysAsnSerSerCysAspGluCysIleArgT 88
|||||
202 TGTGGTCATTCCTGCTGCTGGAACAGTTACTGTGATGATGATATAGAA 251
88 hThrLeuLeuGluSerAspLysHisThrCysProThrCysHisGlnAsn 104
|||||
252 CAGCACTCTGGAGTCAGATGAAACATACATATGTCACATGTCACAAAT 301
105 AspValSerProAspAlaLeuIleAlaAsnLysPheLeuArgGlnAla 121
|||||
302 GAGTTTCCTCGATGCTTAAATGCAACAAAGTTTACGACAGCTGT 351
121 LysAsnPheLysAsnGluThrGlyTyrThrLysArgLeuArgLysGlu 138
|||||
352 TAATTAATTTAAATGAAACTGCTATATCAAAACAGCTACGAAACAGT 401
138 euProProHeuLeuPheLeuValProProArgProLeuSerGlnArg 154
|||||
402 TACCGCACCCACCCACCCAGTACACCAACAGACACATCATGACGAG 451
155 AsnLeuGlnProArgSerArgSerProIleLeuArgGlnGlnAspPro 171
|||||
452 AACCTGCAGCTCTGATGATGATCTCCATATCATCAAGACAGAGATCT 501
171 ValPheArgTyrThrValSerProThrCysSerAspThrLysThrAla 188
|||||
502 GATGATTCAGTGACATGCTCTCCGCTCACTCA..... 535
188 LysCysSerAspSerGlyThrLeuSerArgLeuProAlaProSerIle 204
|||||
536GCTCCCTCATTA 547
205 SerSerLeuThrSerAsnGlnSerSerLeuAlaProProValSerGly 221
|||||
548 TCTTCACTTAATCAATTCATCTGCCCTGCTCTTCTGCTGGA 597
221 nProSerSerAlaProAlaProValProAspIleThrAlaThrValSer 238
|||||
598 TCCATTCTTGTGCCCCAGCTCCAGTACCTGATATACCTGCAACCGTGT 647
238 LeSerValHisSerGluLysSerAspGlyProPheArgAspSerAspAsn 254
|||||
648 TATCATCCCACTCAGAAAAATCGATGAGACTTTGGGATTTCTGATAT 697
255 LysLeuLeuProAlaAlaAlaLeuThrSerGluHisSerLysGlyAla 271
|||||
698 AATTAATTTGCGACCTGCGCCCTTACATCAAGAACATTTCAAAGGAGC 747
271 rSerIleAlaIleThrAlaLeuMetGluLysGlyValProGlyThrS 288

|||||
748 TTCAAATGCTATTACGCTCTTATGSAAGAAAGGGCTA.CCAGGTACCA 796
288 erProTrpAsnSerIlePheVal.GlyGlnSerLeuLeuHisGlyLne 304
|||||
797 GTCCCTTGGAACATCCATCTTTGTGGACAGTCATATATACATGAGACGT 846
304 utLeProThrThrGlyProValArgIleAsnAlaAlaArgProGlyG 321
|||||
847 GATTCACACACTGGGCCAGTAAGATCAATGCTGCTCCAGAGGTGTG 896
321 LysArgProGlyTrpGluHisSerAsnLysLeuGlyTyrLeuValSerPro 337
|||||
897 GCGGCGACAGCTGGGAGCATTCACAAACACTGGGATCTAGTTCTCCA 946
338 ProGlnGlnIleArgArgGlyGluArgSerCysTyrArgSerIleAsn 354
|||||
947 CCACAGCAAAATTAGAAAGAGAGAAAGAACTGTTACAGAAATATAAAC 996
354 gGlyArgHisHisSerGluArgSerGlnArgThrGlnSerProSerLeu 371
|||||
997 CCGGCGACACACAGCGAAGATCACAGAGACTCAAGGCCATCACTTC 1046
371 roAlaThrProCysPheValProValProProProLeuTyrProPro 387
|||||
1047 CAGCACTCAGCTCTTGTGCGGTTCCACACACTCTTGTATGCGGCT 1096
388 ProProHisThrLeuProLeuProProGlyValProProGlnPheSe 404
|||||
1097 CCCCCCATACACTTCTCTCTCCAGGTGTAACCTCCCCACAGTTTC 1146
404 rProGlnPheProSerSerGlnProProThrAlaGlyTyrSerValPro 421
|||||
1147 TCCTCAGTTTCTCTGCGCACGCTCCACACGAGGATATAGTCTCTC 1196
421 roProGlyPheProProAlaProAlaAsnIleSerThrLacyPheSer 437
|||||
1197 CTCACAGGTTTCCACAGCTCTCGCCAATATATCAACACTTGGATCA 1246
438 ProGlyValProThrAlaHisSerAsnThrMetProThrThrGlnAla 454
|||||
1247 TCAGAGATGACAGCTCCCATTCAAATACCATCCCTACAAACACAC 1296
454 oleuLeuSerArgGluGluPheTyrArgGluGln..... 465
1297 TCCTTGTCCAGGGAAGAAATTCTATAGAGACAGACGCTTAAGAA 1346
466AsnAspLysGlyArg 470
1347 AGCAAAAGAAAAAGTCCAAAGCTAGATGAGTTTCAAAATGATTTTGC 1396
471 Glu..... 471
1397 GAATGTATGGAATACAAAAAGATTCAAAAGAGAGCTAGGCGCTCATTT 1446
472SerLysPheProTyrSerGlySerTyrLysArgSerSerTyr 487
1447 CAGGTATAATCTCCCTATAGTGGGATCGATGTAAGAAGTTCTATCA 1496
487 hrAspSer.SerGlnGlyLeuAlaGlnHisIleHisAlaLeuThrLeu 503
|||||
1497 CTATATTCMAAGTCAAGTCTGCGCTCAACAGCTTCAGAGCTTTACT 1546
503 rProSerAlaAlaHisThrLeuAspLeuLeuHisAspHisProHisPro 520
|||||
1547 TCTTACGCGCTCACATTTCTCGTCTTACAGATCAACCCCAATACCC 1596
520 roGluGluAlaGluAlaArgSerAlaMetLysValHisMetProAspLeu 536
|||||
1597 CAGGAGAGCGAGAGCAAGAGCCGCAATATTCGTTACAGGTCACAGATTC 1646
537 MetAspIleAlaHisAlaArgSerArgSerProProTyrArgTyrArg 553
|||||

1647 ACGGATACAC .CGATCTAGGTCAAGTCACCTCCCTATAGACGATACA 1695
553 gSerArgSerArgSerProGluPheArgGlyGlnSerProThrLysA 570
1696 CTCACGGTCCAGATCTCTCAAGCATTTAGGGGACACTCTCCCTAAC 1745
570 rGAsnValProArgGluGlyGlnArgGluTyrPheAsnArgTyrArg 586
1746 GCATGTACCTCAAGAGAAACAGCGGTGAGTATTATATAGATACAGA 1795
587 GluValProProProTyrAspIleLysAlaTyrTyrGlyArgSerValAs 603
1796 GAAGTTCACCCCTTATGACATCAAGCCATATATGGCGGAGTGTGCA 1845
603 pPheArgAspProPheGluLysGluArgTyrArgGluTyrPgluArgLysT 620
1846 CTTTAGAGACCCATTGTGAAAGAACGCTACGGGAATGGGAAAGGAAT 1895
620 yTArgGluTyrTyrGluLysTyrTyrLysGlyTyrAlaValGlyAlaGln 636
1896 ACCGAGAGTGTATGAGAGTACTACAAAGGGTACGCGGTGGAGCTCAA 1945
637 ProArgProSerAlaAsnArgGluAspPheSerProGluArgLeuLeuPr 653
1946 CTTAGACCTCAGCCAAATAGAGAGACTTTTCTCAGAGAGACTCTTACC 1995
653 OLeuAsnIleArgAsnSerProPheThrArgGlyArgArgGluAspTyrA 670
1996 TCTTAATATCAGAAATTCACCCCTTACAGAGGCGGAGAGACGATAG 2045
670 IAlaIleGlnSerHisArgAsnArgAsnLeuGlyLysAsnTyrProGlu 686
2046 CTGCTGGCAAGTCATAGAAATAGAAATCTAGGTGGCACTATCCAGAA 2095
687 LysLeuSerThrArgAspSerHisAsnAlaLysAspAsnProLysSerLys 703
2096 AAGCTTCAACAGAGGACAGTCACATGCAAAAGATATCCAAATCGAA 2145
703 sGluLysGluSerGluAsnValProGlyAspGlyLysGlyAsnLysHisL 720
2146 GGAGAGAGAGAGTGAATGTTCAGAGAGCGCAAGGAAAGCAAGCAT 2195
720 ySLysHisArgLysArgArgAsnGluGlyLysGlyGluLysSerGluSer 736
2196 AGAAACAGAGAAACGAGA AAGGGGAGAGAGAGTACAGAGC 2236
737 PheLeuAsnProGluLeuLeuGluThrSerArgLysCysArgGlySerSe 753
2237 TTCTGAAACCCAGACCTACTGGAGACGTCTAGGAATGCAGGGAATCGTC 2286
753 rGlyIleAspGluThrLysThrAspThrLeuPheValLeuProSerArgA 770
2287 AGGATTCATGAACGAAAGACAGATACACTGTTGTTCTCTCCAGACAG 2336
770 sPAspAlaThrProValArgAspGluProMetAspAlaGluSerIleThr 786
2337 ACGATGCTACACCTGTAGGATGAGCCAAATGAGCGCAAGATCGATCCT 2386
787 PheLysSerValSerAspLysAspLysArgGluLysAspLysProLysVa 803
2387 TTCAGATCAGATCTGACAAAGACAGAGGAAAGGATTAACCCCAAAAGT 2436
803 LysSerAspLysThrLysArgLysSerAspGlySerAlaThrAlaLysL 820
2437 AAAAAGTACAAAGCAAGGAAAGTGAAGGCTCTCTACAGCAAGA 2486
820 ySAspAsnValLeuLysProSerLysGlyProGlnGluLysValAspGly 836
2487 AAGACAATGTTTAAACCTCTTAAGGACCTCAAGAAAGATAGATGGA 2536
837 AspArgGluLysSerProArgSerGluProProLeuLysLysAlaLysG 853
2537 GACCGTGAAGTCTCTCTGCTGAGCGCCACTCAAAAAGGCCAAAGA 2586

853 uGluAlaThrLysIleAspSerValLysProSerSerSerGlnLysA 870
2587 GGAGGCTACAAAGATGACTCTGTAACCTCTCCTCTTCACAGAG 2636
870 sPgluLysValThrGlyThrProArgLysAlaHisSerLysSerAlaLys 886
2637 ATGGAAGGTCACCTGGAACCCCTAGAAAGCCCATTTCTTAATCTGCAAAA 2686
887 Asp . ThrArgArgGlnSerGlnProArgThrArgArgSerLysArgThrV 903
2687 GAACACCAAGAGGCAAAAGCCAGCAAGAGAGAGATCAAAAAGGACTG 2736
903 aLProLysThrSerSerGlnLysSerGlnProValArgThrArgArgPro 919
2737 TTCAAAGACATCAAGTCAGAAAGCCAGTACAGAGAGAGAGGCCCA 2786
920 ArgSerLeuArgLysIleAsnTyrLeuIleAlaArgGluLysAsnGluArg 936
2787 AGAAGCCTGAGAAATTAACCTACTGATGCAAGAGAGAGAAACGAAG 2836
936 gGluLysArgLys . LysSerValAspLysAspPheGluSerSerSerMe 952
2837 AGAAAGCGAGAGAAAGAGTGTAGATTAAGATTTGAGTCTTCGAAT 2886
952 tLysIleSerLysValGluGlyThrGluIleValLysProSerProLysA 969
2887 GAAATCTCTTAAGTAGAAGGACAGCAAAATAGTAAACCATCACCAAAAC 2936
969 rGlyLysMetGluGlyAspValGluLysLeuGluArgThrProGluLysAsp 985
2937 GGAATAGGAAAGGTGATGTTGAAAGCTCGAAAGGACCCCGAAAGGAC 2986
986 LysIleAlaSerSerThrThrProAlaLysLysIleLysLeuAsnArgG 1002
2987 AAGATTCGATCATCACTACCTCCAGCCAAATAATCAAACTCAACAGAG 3036
1002 uThrGlyLysLysIleGlyAsnAlaGluAsnAlaSerThrThrLysGluP 1019
3037 AACCTGAAAGAAATTTGGAAATGCAAGAAATGCAATCTACACAAAGAAC 3086
1019 rOserGluLysLeuGluSerThrSerSerLysIleLysGlnGluLysVal 1035
3087 CCTCTGAAATTTGGAGTCAACATCTACCAAAATCAAAACAGAAAGATC 3136
1036 LysGlyLysAlaLysArgLysValAlaGlySerGluGlySerSerSerTh 1052
3137 AAGGAAAGGCCCAAGGAAAGTAGCTGGTGGGAGGCTCCACCTCCAC 3186
1052 rLeuValAspTyrThrSerThrSerSerThrGlyLysSerProValArgL 1069
3187 GCTGTGATTTACACAGTACAACTTCAACTGAGGACAGTCTGTGAGA 3236
1069 ySerSerGluGluLysThrAspThrLysArgThrValIleLysThrMetGlu 1085
3237 AATCTGAAGAAAGACAGATCAAAAGCAAGACAGTCATTAAACATATGAG 3286
1086 GluTyrAsnAsnAspAsnThrAlaProAlaGluAspValIleIleMetI 1102
3287 GAATATATATATGATTAACACAGCTCTCTGACAGATGTTAATAATATGAT 3336
1102 eGluValProGlnSerLysTrpAspLysAspAspPheGluSerGluGluG 1119
3337 CCAGGTTCTCTCAGTCAAAATGGAATAAGATGACTTGAAGTCAAGAGAG 3386
1119 LuAspValLysThrThrGlnProIleGlnSerValGlyLysProSerSer 1135
3387 AAGATGTTAAACCAACACAACTATACAGAGTGTAGGAAACCAATCGAGT 3436
1136 IleIleLysAsnValThrThrLysProSerAlaThrAlaLysTyrThrG 1152
3437 ATTATAAAAATGTCACCTACTAAGCCATGCGCTACGGCTAAGTACACCGA 3486

1152	ulysglusertgluglnproglulysleuulinlysleuprollys(lu)alas	1169
3487	GAAGGAACCGACAGACCCGAGAACTGGAGAACTCCCAAGGAGGCCA	3536
1169	erh1sgluleumetglnh1sgluleuArgSerSerlysglySerAlaser	1185
3537	GGCAGAGAGTGTATGCAGCAGCAGACTCAGGAGCTCAAGGGCAGTGGCTCC	3586
1186	SerGluLysGlyArgAlaLysAspArgLuh1SerGlySerGluLysAs	1202
3587	ACTGGAAGGCGCAGAGCCAAAGACCAGGAGCTCAGAGGCTCGAGAGAAGA	3636
1202	pasnProAspLysArgLysSerGlyAlaGlnProAspLysGluSerThy	1219
3637	CAACCTGTCAAGAGAGAAAGCGGTGCCAGCCAGCAAGAGAGAGACAG	3686
1219	alAspArgLeuSerGluGlnGlnh1sphaLysThrLeuSerGlnSerSer	1235
3687	TGGACCGCTGTGATGACAGGACATTTTAAACTCTCTCTCAAGCTCTCC	3736
1236	LysGluThrArgThrSerGluLysH1sgluserValArgGlySerSerAs	1252
3737	AAAGAGACCAAGACATTCAGAGAAAGCAGAGTCTGTTCTGGTCTCCCA	3786
1252	nLysAspPheThrProGlyArgAspLysLysValAspTyrAspSerArg	1269
3787	TAAAGACTTCACTCTCTGTAGAGACAAAGAAATGGACTAGCAGACAGGG	3836
1269	sPTyrSerSerSerLysArgArgAspGluArgGlyGluLeuAlaArgAla	1285
3837	ATTATTCCAGTTCACAAAGCAAGACGAAAGAGGTGAATTACCAAGGAGA	3886
1286	LysAspSerProProArgGlyLysGluSerLeuSerGlyGlyLysSerly	1302
3887	AAAGACTCTCTCCCGGGGCAAAAGTCTCTGTCTGGGCAAAAAAGCAA	3936
1302	sLeuArgGluLysArgAspLeuProLysLysGlyAlaGluSerLysLys	1319
3937	GCTGAGGCGAGAGAGATTACTTAAAGGGGGCGAGTCAAAAAA	3986
1319	erAsnSerSerProProArgAspLysLysProH1AspH1sLysAlaPro	1335
3987	GTAATTCTAGCCCCCAAGAGACAAAACCTCATGTATGATAAGGCCCC	4036
1336	TyrGluThrLysArgProCysGluGluThrLysProValAspLysAsnSe	1352
4037	TACGAACCTAAACGCCCATGTGACAGACAAAAGCCTGTGATTAATAATCC	4086
1352	rgLysGluArgGluLysH1sAlaAlaGluAlaArgAsnGlyLysGluS	1369
4087	TGGGAAGGAGCGGAGAAAGCATGCTCTCAACTGCCAATGGGAAGAAT	4136
1369	erSerGlyAlaAsnCysH1sValTyrLeuThrArgGlnThrLeuProTr	1385
4137	CCAGTGTGGCAAACTGCATGTATCCCTAACCCGCAAGACCTCCCATG	4186
1385	pArgArgSerTrpLeuLeuGlyArgThrPArgArgAlaProSerSerArg	1402
4187	GGAAGAGACTGGCTGTGGCAAGGTGGAGAAAGGCGCTCAAGCGGAA	4236
1402	snProSer	1404
4237	ACCCGAGC	4244

seq_name: gb_pr:AF352051

seq_documentation_block:

LOCUS AF352051

DEFINITION Homo sapiens proliferation potential-related protein mRNA, complete cds.

ACCESSION AF352051

VERSION AF352051.1

KEYWORDS GI:15705402

5376 bp

mRNA

linear

PRI 20-SEP-2001

Source	Organism	Human
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
AUTHORS	Camargo, A.A., Moreira, E.S. and Simpson, A.J.G.	
JOURNAL	Direct Submission	
TITLE	Submitted (21-FEB-2001) Laboratory of Cancer Genetics, Ludwig Institute for Cancer Research, Rua Antonio Prudente 109 4th floor, Sao Paulo 01509-010, Brazil	
FEATURES	Location/Qualifiers	
source	1..5376	
	/organism="Homo sapiens"	
CDS	/db_xref="taxon:9606"	
	92..4942	
	/note="Similar to mouse proliferation potential-related protein"	
	/codon_start=1	
	/product="proliferation potential-related protein"	
	/protein_id="AL05625.1"	
	/db_xref="GI:15705403"	
	/translation="MGIKTLNLVGLKRLAEFPEVFMMEVQPNMKGAMLNIGKVAI PTIDAEVATGKKKKPPPLPEEPSSSEDDPIPELLCLICIDIMDAVIVCCGSS YCDCTIRALLESDEHNCPTCHODVSDPDIANKPIQAVNNFRKNTGTYKRLKRL PDPPIPIPPPLIQRNLQPLMRSPISRQDDPLIPVTSSTTHPAPISLTSKQSL APVSGNPSAPAVPDTITAVISLVSHEKSDGDFRSDMKILPAAALAEHSGKSS IALTALMEKGYQVPLGTESILGOSLIHQILPTGPVINTPARGGRPMGEHSHK LGYVSPPOIRRGESRCYRSINGRHSHSEKSTOGSLPAPVEVPVPPPLYPPI PHPIPLPGVPPPPGPPGPPGPPAGYVPPPGPPAPANLSTPVSSGCVTAHNN TIPPTQAPLPSREFTYRQRLKTEEKSKLDEFTNDPAKELMETKIKTORERRSS RSKPSGSGSYSSRSYTSKRSQSTSRYSRFSHRSRSYSPRYRGKRSR NYRBSRSHGHYHRSRSRSPYRHYHRSRSPOARFGSPKRNVPQDETRFYFNRYR EVPPPYDMKAYIGRSDPDEFEKEREMERKRYEYKAYAGAAPRSANE NESERETPLINRSPPTGRREDYVGGSHRSNTISNYPEKLSADGNKDNKTS KEKSEANPDGDKGNKKKKKKRKGESGEFLNPELLHSRSREPTGYEENKTSGL FVLSDRDATVPDEPDMAESTITKSYSEKDKRDEPKFKAAGDTTKRNGSAVSKKE NIVAPAGPOEYGVGERKSRSPRIKKAAETPKPDNTRSSSSOKDERIGTGPPE AHSASAKHOETPKVKEKKYKADYSDVSKSLTKTEKAKAPNEKKPLDNGEKRR RKTEGEVNDKDFESSMKISKELEVTEYKSPKMPDEKMDRPEKDISLAPAK KKILNRETGKIGISTENISNTKEPSELESTSSKOEYKGVKRVKIVGTGESSST LVDYTSSTGSGVPRKSEKTDPRKTVITKMEYVNDNDNAPADYTIMIYVQSKMD KDPSEBEDYKSTOPISSYKGRASVYKNNSTKSNLYKPREKESPESEKIQKTRKY SHELTIOHEVSSKNSASSEKGTIDROYVLEKSNPKRKNSTOPEKESMLDRNDEGG NEFLSQSKSEKARTSDKHDRASNNKDPENRDKNDYDREYSSSKRDEKNEFLR KCDPSRKNDSASGOKNKPREDRLPKKGDSGKSSNSRDKRPHDKATYDTKRP NEEKRSYDNKPCDKREKHLVLEARNKKEGSKKILYINLPETOYEKQITGQIDKSV KPKQOLSHSRSLSDLTRREDEAPFEPDYNESDESNAVYKEEESSNISKDLKDKIV EKKAESLDTAAVVOGILSRNDSHSPSRHSRPSGSDTRSHSSASAEQSDSKR KKKKKKKKKKRKKRKKRKKHAKHAELEVELERKSOHKKRKKRKKKDEKEKEDDK VKSYTV"	
BASE COUNT	2008 a 1082 c 1074 g 1212 t	
ORIGIN		
alignment_scores:	Quality: 5739.00	Length: 1445
	Ratio: 4.477	Gaps: 12
Percent Similarity:	88.720	Percent Identity: 78.062
alignment_block:	US-09-811-045A-1 x AF352051	..
Align seg 1/1 to: AF352051 from: 1 to: 5376		
1	MeMeGluValIlylsasprProksnmctlysglyAlaMelLeuthrasmh 17	
158	ATGATGGAAGTGAAGATCCTAATATGAAAGGCGCATCTTACCAACAC 207	
17	rglyLyTyYrAlaIleProThrIleasplaagIuaIaTyYrAlaIleGly 34	
208	TGCAAAATATGCATATCCAACTATATGATGACAGACATATGCATTGGGA 257	
34	yslysgIuIySprProPhleuProGluGluProSerSerSerSerGlu 50	

```

1103 CTGGTTTCCACACCAACAAATTAGAAAGAGGGAGAGGAGCTGCTACAG 1152
350 gserlleasnargylarqghishlssegiarqsergiarqthrglins 367
1153 AAGTATAAACCGTGGGCGACACACACGCGAAGAAAGATCACAGAGACTCAAG 1202
367 erProSerLeuProAlaThrProCysPheValProValProProPro 383
1203 GCCCGCTACTACCAGCACTCCAGCTTGTGGACGTTGCACGACCTGCT 1252
384 LeuTyrProProProProHisThrLeuProLeuProProGlyValProPr 400
1253 TTGTATCCGCTCCCTCCCATCATCTCTCTCCCTCCGGGTGTCTCTCC 1302
400 oProGlnPheSerProGlnPheProSerSerSerProProThrAlaLyt 417
1303 TTCACAGTTTCTCCCACTTTCCTTCCTCGGCGACGACACCCTGGGT 1352
417 yfSerValProProProGlyPheProProAlaProAlaasnIleSerThr 433
1353 ATATGTGCCCTCTCCAGGGTTTCGTCGACGCTCGGCAATTATTCACAA 1402
434 AlaCysPheSerProGlyValProThrAlaHisSerAsnThrMetProth 450
1403 CCTGGGTGATCAACAGAGAGTGCAGACGTCAATTAAATACATCCCAAC 1452
450 rThrGlnAlaProLeuLeuSerArgGluGluPheTyrArgGluGln... 465
1453 AACCAAGACACCACTTGTCCAGGAGAAATTCATATAGAGACGACGAC 1502
466 .....Asn 466
1503 GACATAAAGAGAGGAAAGAAAGATCCACACTGATGACTTACAAAT 1552
467 AspLysGlyArgGlu..... 471
1553 GATTTTCTAAGAAATGTATGCAATCAAAAAAGATTCAAAAGAGCCGTAG 1602
472 .....SerLysPheProTyrSerGlySerTyrSerA 483
1603 GCGCCTATTTCCAGGCTATAAATCTCCCTATAGTGGTCTCTGATTCTAA 1652
483 rgSerSerTyrThrAspSer.SerGlnGlyLeuAlaGlnHisIleHisAl 499
1653 GAACTTATATTAATTATTCAAATCAAGATCGGGCAACAGTTCCACGC 1702
499 aleuThrLeuSerProSerAlaAlaHisThrLeuAspLeuHisAspH 516
1703 TCTTAATCTCGATCATTCACGCCCTCACATTCCTGCTTCAATTCACGGT 1752
516 lsrProHisProProGluGluAlaGluAlaArgSerAlaMetIleHisAl 532
1753 ACCCTCCATACCCTCAGGAAGGCGAGGCAAGAGCCCAATTAATCCGTAC 1802
533 MetProAspLeuMetAspIleAlaHisAlaArgSerArgSerProProTyr 549
1803 GGTCTAGTCTCATGATATCAT..CCCCCCCCCCCCCTTA 1853
549 raArgArgTyrArgSerArgSerArgSerProProGluPheArgGlyGln 566
1852 CAGACGCTATCATTCACGATCAAGATCTCTCAAGCGTTAGGGGACAGT 1901
566 erProThrLysArgAsnValProArgGluGluLysGluArgGluTyrPhe 582
1902 CTCCCAATAAACGTAAATGTACTCTCAAGGGGCAACACACAGTAAATATTT 1951
583 AsnArgTyrArgGluValProProProTyrAspIleLysAlaTyrTyrGln 599
1952 AATGATATCAAGAGAAGTTCACACACCTATGACATATAAGCATATATTAG 2001
599 yArSerValAspPheArgAspProPheGluLysGluArgTyrArgGluT 616
2002 GAGAGAGTGTACTTATGAGACCCATTTTGAAAAAGAACGCTACCGAGAT 2051

```

616	rfGluuArgLysTyrArgGluuTfprLyGluLysTyrTyrLyGlyTyrAla	632
2052	GGAGAGGAAAAATATGAGAGTGGTATGAAAAATATATATAAAGGTTATGCT	2101
633	ValGlyAlaGluProArgProSerAlaAsnArgGluAspPheSerProG	649
2102	GCTGGAGCACACCTTAGACCTTCACCAATATGAGGAACCTTTCTCCAGA	2151
649	uArgLeuLeuProLeuAsnIleArgAsnSerProPheThrArgLysArg	666
2152	GAGATTTTGGCATTACATCAGAAATTCTCCCTTCACAGAGGCCGCA	2201
666	rgGluAspTyrTyrAlaGlyGlnSerHisArgAsnArgAsnLeuGlyGly	682
2202	GAGAAAGACTATTTGGTGGCCAAATCTATAGAAAGTCGAAACATGAGTAC	2251
683	AAsnTyrProGluLysLeuSerThrArgAspSerHisAsnAlaLysAspAs	699
2252	AACTATCCGAAAGAACCTTTCACAGCAAGATGTGTCAATTCAGAAAGATA	2301
699	nProLysSerLysGluLysGluSerGluAsnValProGlyAspGlyLysG	716
2302	TTCAAAGTCAAAAAGAGAGAGAGTGAAGAAAGCTCCAGAGATGGTAAAG	2351
716	LysAsnLysHisLysLysHisArgLysArgLysArgAsnGluGluGlyGlu	732
2352	GAAATATAGCATATAGAAACACAGAAAAAGAAAGH.....AAAGGGAG	2392
733	GluSerGluSerPheLeuAsnProGluLeuLeuGlnThrSerArgLysCys	749
2393	GAAAGGAGAGCGTTTCTGTGACCCAGAGATTATTAGAGACTTGTAGAAATC	2442
749	sArgGlySerSerGlyIleAspGluThrLysThrAspThrLeuPheVal	766
2443	AAAGAGAACTACAGGTTGTAAGAAAAATTAACACAGCATATGTTGTCTC	2492
766	eupProSerArgAspAspAlaThrProValArgAspGluProMetAspAla	782
2493	TCCCAAGTGAAGATGATGGCACACCTGTTAGAGTAGAACCAATGGATGCA	2542
783	GluSerIleThrPheLysSerValSerAspLysAspLysArgGluLysAs	799
2543	GAATCATCTACTTTTAATCACTGTCTCTAAAAAGCAAGAAAGAAAGGA	2592
799	plysProLysValLysSerAspLysThrLysArgLysSerAspGlySerA	816
2593	TAAACCAAAAGCAAAAGGCGATTAACCAAAACGAGAAATGATGATCTG	2642
816	IaThrAlaLysLysAspAsnValLeuLysProSerLysGlyProGlnGlu	832
2643	CGGTGTCCAAAAAAGAAATATGTAAAAACCTGCTAAAGGAGACCCAGAA	2692
833	LysValAspGlyAspArgGluLysSerProArgSerGluProProLeuLys	849
2693	AAAGTATGATGGAAGAGTGAAGATCTCTCGATGTGAACCTCCCAATTAA	2742
849	sluysAlaLysGluGlnAlaThrLysIleAspSerValLysProSerSers	866
2743	AAAAGCCAAAGAGGAGACCTCCCAAGACGTGACAATCTAAATCATCATCTT	2792
866	eTserGlnLysAspGluLysValIleThrGlyThrProArgLysAlaHisSer	882
2793	CCCTTCAGAAAGATGAAAAAAACACTGGAGACCCCCAGAAAAGCTCATCT	2842
883	LysSerAlaLysAsp..ThrArgArgGlnSerGlnProArgThrArgArgS	899
2843	AAATCAGCAAAAGAACACCAAGAAACAAACACGATCAAGAGAAAGAAAGT	2892
899	eLysArgThrValProLysThrSerSerGlnLysSerGlnProValArg	915
8893	GAGAGAGGACTATTTCCAAAGATGTAAATTCAGAAAAAGCTAACATCAAG	2942

916	ThrArgProArgSerLeu...ArgLysIleAsnTryPheIleIleAlaIrr	931
917	932
918	933
919	934
920	935
921	936
922	937
923	938
924	939
925	940
926	941
927	942
928	943
929	944
930	945
931	946
932	947
933	948
934	949
935	950
936	951
937	952
938	953
939	954
940	955
941	956
942	957
943	958
944	959
945	960
946	961
947	962
948	963
949	964
950	965
951	966
952	967
953	968
954	969
955	970
956	971
957	972
958	973
959	974
960	975
961	976
962	977
963	978
964	979
965	980
966	981
967	982
968	983
969	984
970	985
971	986
972	987
973	988
974	989
975	990
976	991
977	992
978	993
979	994
980	995
981	996
982	997
983	998
984	999
985	1000
986	1001
987	1002
988	1003
989	1004
990	1005
991	1006
992	1007
993	1008
994	1009
995	1010
996	1011
997	1012
998	1013
999	1014
1000	1015
1001	1016
1002	1017
1003	1018
1004	1019
1005	1020
1006	1021
1007	1022
1008	1023
1009	1024
1010	1025
1011	1026
1012	1027
1013	1028
1014	1029
1015	1030
1016	1031
1017	1032
1018	1033
1019	1034
1020	1035
1021	1036
1022	1037
1023	1038
1024	1039
1025	1040
1026	1041
1027	1042
1028	1043
1029	1044
1030	1045
1031	1046
1032	1047
1033	1048
1034	1049
1035	1050
1036	1051
1037	1052
1038	1053
1039	1054

3840 AAAAAAAAAAGTAATTTGGACCGTCTGAAGAACAGAAATTTTAAAGCT 3889
1231 laeuSerGlnSerSerIySgluThrArgThrSerGluYsHisGluSerA 1247
3890 CTGTCTCAATCTTCCAAAGAGCGTAGACGTACGATTAACATGATTCAC 3939
1247 lArgGlySerSerAsnIySAspPheThrProGlyArgAspIySValA 1264
3940 TCGTGTCTTCCAAATTAAGACTTCTCCATAGACACAAAAAAACGTG 3989
1264 sPTyTrAspSerArgAspIyTrSerSerSerIyArgArgAspGluArgGly 1280
3990 ACTATGACACACAGAGAGTATTCAGATTCCAACGATGAGATGAAAGAT 4039
1281 GluLeuAlaIArgArgIySAspSerProProArgGlyIySgluSerIySe 1297
4040 GAATTACAGACAGCAAGAAAGACTCTCTTCGGAATTAAGATTCGATC 4089
1297 rGlyGlnIySerIySLeuArgGluGluArgAspLeuProIySlySgly 1313
4090 TGGACAGAAAAATTAACCAAGGAGAGAGATTTGCCCTAAAAAAGGAA 4139
1314 .AlaGluSerIySlySerAsnSerSerProProArgAspIySlySPro 1329
4140 CAGGAGATTCCAAAAAAGTAATTCAGTCCTCCACAGACAGAAAAACCT 4189
1330 HisAspHisIySAlaProIyTrGluThrIySArgProCySgluGluThrIy 1346
4190 CATGATCCAAAAAGCCACTTATGATTAACGCGCAATGAAAGAGACAA 4239
1346 sProValaIySAsnSerGlyIySgluArgIySHisAlaIySAlaGluA 1363
4240 ANCTGTAGATTAATAATCCCTTGAAGATGCTGAGAAAGCATGATTTGA 4289
1363 lArgAsnGlyIySgluSerSerGlyAlaAsnCySHisValIyTrleuTh 1379
4290 CAGGAAACATTAAGAGTCAGAGGCAATTAACCTATTATTACTTAAC 4339
1379 rArgGlnIyTrleuProIyArgArgSerTrIySleuGluIyArgTrIyP 1396
4340 CCACCGAGACACAGGTTGAAAAAGCAATTACTGGCAAAATTGACAA 4389
1396 rGAlaProSerSerArgAsnProSer 1404
4390 GAGTACTGTCAAGCCTAAACCCAGT 4415
seq_name: gb_AC010321
seq_documentation_block:
LOCUS AC010321 207256 bp DNA linear PRI 18-DEC-2001
DEFINITION Homo sapiens chromosome 16 clone CTD-2540M10, complete sequence.
ACCESSION AC010321
VERSION AC010321.8 GI:17921202
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 207256)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 207256)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 207256)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (26-SEP-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA

REFERENCE 4 (bases 1 to 207256)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Dec 18, 2001 this sequence version replaced gi:15778689.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.5% of Sequence;
Estimated Total Number of Errors is 0.4.
FEATURES
source
1. 207256
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="CTD-2540M10"
BASE COUNT 61164 a 43640 c 45330 g 57122 t
ORIGIN
alignment_scores:
Quality: 4135.50 Length: 3462
Ratio: 3.443 Gaps: 23
Percent Similarity: 34.691 Percent Identity: 30.156
alignment_block:
US-09-811-045A-1 x AC010321/rev ..
Align seg 1/1 to reverse of: AC010321 from: 1 to: 207256
86 lIeaGThrThrIleuIeuGluSerAspIySHisTrCySProThrCySHi 102
99095 ATTAAGACAGCACTCTCTGGAATCAGATGACACACATGCGCACGTGTA 99046
102 sGlnAsnAspValSerProAspAlaIleuIleAlaSnIySPhleuArgG 119
99045 TCAAAATGATGTTCTCCTGATGCTTAATTCCAATAAATTTTACGAC 98996
119 lN..... 119
98995 AGTAACTGTCTGTATCCATTTTATGAAAAAATCTTTTAACTGATTTA 98946
119 119
98945 ACTGTACTTCACTGAATCTGCATAACATTTTCTGCATTTATTATGCTTG 98896
120AlaValaSnAsnPhelySAsnGluThrGlyTrThrIySA 133
98895 GATCTGTAGCGCTGTAATAACTCAAAAATGAAATGAACTGCTATACAAAA 98846
133 rGluuArgIySgluIeuProProPhleuPhleuValProProArg 149
98845 GACTACGAAACAGTTACTCTCTCCACACACCCCAATACACCTCGAGAA 98796
150 ProIeuSerGlnArgAsnIeuGluProArgSerArgSerProIleuArg 166
98795 CCACGTGATTCACAGGACACTACACCTCGATGAGATGTCGATTCAG 98746
166 gGlnGlnAspProValValPheArgTyTrThrValSerProThrCySserA 183
98745 ACAACAGATCTCTTATGATTCAGTGACATCACTCAATCACT..... 98703
183 sPTyTrIyThrIaGlySerSerAspSerGlyTrThrIeuSerArgIeu 199
98702CAC 98700
200 ProAlaProSerIleSerSerIyThrSerAsnGlnSerSerIeuAlaPr 216
98699 CCAGCTCGGTATATCTTCACTTAACCTCTATACGTCTCTCTTGCCGCC 98650
216 oProValSerGlyAsnProSerSerAlaProAlaProValProAspIleT 233

|||||
98649 TCCTGTCGGAATCCGCTCTCTCAGCTCCTGACCGATATATA 98600
233 hralaThrValSerIleSerValHisSerGluYsSerAspGlyPro... 248
98599 CTGCAACAGATCCATCATCTAGTTCAATTCAGAAAAATCAGATGAGCCTTTT 98550
248 248
98549 CGGTAAAGCCTGTGTGTTTTCACCTGTAGAAACCAATGAGTCAATGAT 98500
248 248
98499 GTAGTGTGTTGTGTGTTATCATCTTTTACAGCTTTCTTACTGAGCA 98450
248 248
98449 CTGTGCTCAGTAATGTGGATGACTGTAAAGATGACTAGTCTTATTTG 98400
248 248
98399 GGGGAAGAGAAAACAACGTGATTTGATTTGGCAATGAGGATTAAGCAT 98350
248 248
98349 CTTGAGGTCAAAATACTACACAGGTGGCAACAGCCCTATTAAGTTG 98300
248 248
98299 AATGTTCCATATACATACACAGCTCTGAATTAACAGAGTGGAGAAAT 98250
248 248
98249 CACTTGTGATCATATGATGATTTTGATAGAGTGTGAGTGTGAGTGAAC 98200
248 248
98199 TAAGAACAGTTTAGATAGTAGAGAAAGGGGTAAAAACACAGATACT 98150
248 248
98149 AGAAGATCAGGAATAGTATGTTTGGACATAATTAATATAGTAGAG 98100
248 248
98099 AATAGTAAACTCACAGCTACCACTTCAGAAATGAGACTTTTTCAT 98050
248 248
98049 GTAGTTAATGTCTCATTAATCTTTAAAAAAGGTGTAG 98000
248 248
97999 AGGCCATTTTCAGACACACTGAAGCAAAATCTTAGATGAGACCTTTAC 97950
248 248
97949 ATTACATTTAAGAAATCCACTGATCTGTGGAATTTATTGGTTGATA 97900
248 248
97899 GTGGAATAGAGCCAGAGAGTTCTATGGCTGACCAAGATTATCTTTCA 97850
248 248
97849 GTTAGTGTAGACTAGAGCTAAACTTAATCTTAGGGGAT 97800
248 248
97799 TTTTGTCTTCATAGCCAGACTTGTGATTTAAGTTTACTATTCTTGCAA 97750
248 248

97749 CTGATTTCTGTATTAACCTAGATATCTTTTAAAGAGGCTTCATTTTA 97700
248 248
97699 CTTTTCCACTTACTACATATAGAGAACTTTCAAGTTTCAGCCTTCTCT 97650
248 248
97649 CACTTTCCTCAGAGATTCATCATCTTTGACTTAACTATCTTAACCTGC 97600
248 248
97599 AGGTGATGGGTAAACAGCAGCAAACTTTTACTCTTAATTTGTCATTTG 97550
249 Phe..ArgAspSerAsp 254
97549 TGAATAATTTTAAATGATATATAATTAATGTTTAAAGGATTCGATA 97500
254 snlYsIleuLeuProAlaAlaIleuThrSerGluHisSerIysGlyAla 270
97499 ATAAATATTGCGCAGCTGCAGCTCTTGCAATCAGAGCACTAAAGGAGACC 97450
271 SerSerIleAlaIleThrAlaIleuMetGlu... 281
97449 TCCTCAATTCATTAACCGCTTATGGAAGAGAGATAATTTACTGG 97400
281 281
97399 TCGTTAATTTGGGATTTTAAACAGAGGTTTGTGTTTCTGCGTTT 97350
281 281
97349 GTTCTTACCTCTTTTTTTTTTAACTATTTCTGTTATTTG 97300
282 LysGlyValProGlyThrSerProTrpAsnSerIlePhe.V 295
97299 GTTGTCTTAAAGGTTACAGAGTGCCTGTTCTGGAACCCCATCTTTCG 97250
295 aLGIyGlnSerIleuLeuHisGlyGlnIleuIleProHyr... 308
97249 TTGGACAGTCATTAATTCATGACAGTGTATCCCAACTGGTGAATA 97200
308 308
97199 GATCAGTTGGTTAGACCTTTCCCTTAAAAAATTTTACGTTGA 97150
309 G 309
97149 TTTAATGTCAATTTTGTATTTACTACCTGCTTTTATTTTGTGTTCTAG 97100
309 lYProValArgIleAsnAlaIleArgProGlyGlyIleArgProGlyTrp 325
97099 GTCCAGTAGAATAAATACGTCGTCCAGGTGTGTCACCAAGGCTGG 97050
326 Glu... 326
97049 GAACAGTGAATGATGTTTACAATAATCTTCAGATGATTCCTGCAGAACT 97000
326 326
96999 AGATAATGGAAGTCCAAACTAGCAAACTAGATAATGGAAGTCCAAAC 96950
326 326
96949 TAGGACGGGTGCTGCTTTTATGTTAACTTCATGTGCCAGATTA 96900
326 326
96899 CATTTTATGTAATATGCAAAATTCAGCTTATAGAGGTTTTTTT 96850
326 326
96849 AAATGAGCAAAATCTTTTCTAAATAACAAATTTTAAAGTTTTTTTCTAC 96800

```

326 ..... 326
96799 TTTCATATCCACAATGGGATCTTGACAGACTTCAGTTAATGATA 96750
326 ..... 326
96749 GATTAGAAATTACTGAAAAGTTATTTAAATATGTCGAGTTATTTCC 96700
326 ..... 326
96699 TAGTAGTTAACTTTTCTACCTTAATTTGGATGATCAAGTTTGATT 96650
326 ..... 326
96649 TGAGTACATAAATACATAATAGATTACGCTACCAACAAAATTTT 96600
326 ..... 326
96599 TGTCTGAATCGAGTGTCTTGCCACCTAAATACAAAGAACTTTGAT 96550
326 ..... 326
96549 TTTCCTGTGGGATATTTTATGCAATTAATAATCAGACACTTACTTAAT 96500
326 ..... 326
96499 GTATTTTACTATATATTAATTTCTGTTATATATCCGATGAAAT 96450
326 ..... 326
96449 GATATATTTCTTAAAGTAGTAGTACTAGCAGATGATCAGTAGAAGA 96400
326 ..... 326
96399 ACATAGATTAGAAACAGTAGACTTGGTTGAATCCCATCTCTTCA 96350
326 ..... 326
96349 TTTCCTAGCTCATGACAGTGTCTGTGTTTACTCCCTAACATCCAT 96300
326 ..... 326
96299 TTCCCTGTGATTTGGAATGGAAAGACACTACCTTAACATGTAATAATGA 96250
326 ..... 326
96249 CTTCATGTGCCAGCAGACAAAAGGCTTTTCTGTATATACAAAT 96200
326 ..... 326
96199 AAGTTCACAGATAATTATAGAGTGAATTAATAATATATATACATA 96150
326 ..... 326
96149 AAGTAATCTAATTTGAGGTCTAGTTATTTGTATGATACAGTTAATGTG 96100
326 ..... 326
96099 GGGGAAATGACTTTAAGTAATAATATATTTACTCAGCTCTGTATCCG 96050
326 ..... 326
96049 AAACAAATTTAAGACTGGGAGTAGAAAAAGTTTCAGAGTAGTTAAA 96000
326 ..... 326
95999 ATGTGTGCTTAAACAGCTATGCAAGAAAGTCAATTAATACCAGG 95950
327 ..... His.SerAsnLysLeu 331
95949 AGAAATACTGAGCCTATTTTCTTCTCCACAGAGTTCCACAAACTT 95900

```

```

332 GlyTyrLeuValSerProProGlnGlnIleArgArgGlyLysSerCys 348
|||||
95899 GCGTATCTGGTTTCTCCACCAACAATTAGAAGGGAGAGAGCTG 95850
348 STYR..... 349
|||||
95849 CTACAGGTAGGCATGCTAAAAATCTGGAAAAATATCATCTATTTT 95800
349 ..... 349
95799 TATATATATCTCACAAAGAAATTAAGCACAGCAGTGCAGAGAGAG 95750
349 ..... 349
95749 GGGGTCAATTTGTTTATGTTGCTTATCCCTTAAGTCTGGGTAGTTG 95700
349 ..... 349
95699 TCCTTAAAGAGGCTAGAAAAGATAATATATATATATATATATGA 95650
349 ..... 349
95649 TGGTTTAAACATTTTATATATCTCTCCAGTGATTAATTTCTCCCT 95600
349 ..... 349
95599 GTCGCAAAACAGACAGATTTGGCTTGCAATTAACCTTGGAAGTAGAA 95550
349 ..... 349
95549 GAAAGTCAGGGGTACAGACATATCTAGATACAGTTAATTAAGATTA 95500
349 ..... 349
95499 AAGTTAACAAAGCTCAGCTCCCACTGTGAACAAAGAAATTCATTTCT 95450
349 ..... 349
95449 CTTCAGACCCCACTGTCCCAATCATCTCCACCGCTCTTTTCCAAAG 95400
349 ..... 349
95399 AAGAGACAGAGATTGTTAGAGCTCCTGCCCTAGTGGCGGATACAAGC 95350
349 ..... 349
95349 AATATTAATCTAAGAAATGAAGATCTTACTGTAGTCAAGGGTCCCA 95300
349 ..... 349
95299 ACCCTGGGCCACAGACCAGTATCCATCCTTGCCGTTAGGAACGTGGCA 95250
349 ..... 349
95249 ACACAGCAGAGGTGAACAGCAGCTGAGCAGATTAACCACTGAGTCC 95200
349 ..... 349
95199 ACCCTGTGTAGATCAGCAGGTGCAATTAAGATCTCATAGAGCATGAAC 95150
349 ..... 349
95149 CCTGTGTGAACCTGCGTGTGCAAAAGATCTGGGTTGCACGCTTCTAGGA 95100
349 ..... 349
95099 GAATCTAATGCCCTGATCATCTAAGGTGAACAGTTTACCCTGCGCC 95050
349 ..... 349
95049 CCGTCCATGAAAAATTTGTTCCACAAACAGCTCCTGTGTCCAAAC 95000
349 ..... 349

```

94999 GGTGGAAGTCTAGTCTAGTCAATATCACACGAGTAAAGAGCC 94950
349 349
94949 TCCAAAGCTTCAATACAGAGTTTGGCCAGCAATAGCTCAGCTGG 94900
349 349
94899 TAATTTAGAGCTTTGGAGGCTGATGGGGAGATGCTTAGGCGCTGA 94850
349 349
94849 AGTCAAGATTAGCTGGGCAACATAGCAAGACTGTCTACCAAAAA 94800
349 349
94799 AATAAAATACAGACAGTGGGATGCACGTGTGCTCCAGCTACCTGG 94750
349 349
94749 GAGGCTGAGCGAGAGATCACTTAGCTTAGGAAGTCAAGCTGCAGTG 94700
349 349
94699 AGCTAATACCATCATCTGCACTTTAGCCTAGGGGACAGAGTGAACCT 94650
349 349
94649 GTCTCCAAAAAGAGAAATGCGGATATCTTAGTGTCTCATGTCTAA 94600
349 349
94599 ACTTATGAAGTTTACTAGTACTTACTGATCAGATCATTTTAAAGAGGT 94550
349 349
94549 AATAGTATTATGCGCTCAGTGAAGATGCTTAAATCTAAACCCCTAC 94500
349 349
94499 TGTGTGTTGTGTACCAATATTATTAATTAAGACCTCTAGTTTAA 94450
349 349
94449 CAGAAATACCAAGTCTGTTGGTGAGCAAAATAGAACTTAGCCAA 94400
349 349
94399 CACTGTATTTTTACTTAAGCTTTTGAGAGAAAGCTTGATATTTTCAT 94350
349 349
94349 GTTAATTATAGAGTTAAACTAGATTATTTCTGCTTGTATGATCTTT 94300
349 349
94299 ACCCACTTCTCCCTCAAAAAAGAGTTACAGAGTTGTTTCTCTCAT 94250
349 349
94249 TTAACCTTTTGGGGCAGGTGTGTGCTCTGACCTGTAAACCCAGCACT 94200
349 349
94199 TTGTGAGATCACTGAGGTGAGAGTTTGAAGACAGCTGCCCAACATG 94150
349 349
94149 GTGAACCTGTCTCTTAAATAACAAAAATTAGCTGGCATGGTGGC 94100
349 349

94099 ACAGCTCTAGTCCAGCTACTTGGGAAGCTGAGCAGAGAAATTCCT 94050
349 349
94049 TGAACCCAGAGGTGAGGTGCAGTGAACCAAGATTACACACTGTACT 94000
349 349
93999 CCAGCCTGGACGACAGAGTGAGACTCTGTCAAAAAATATAAATAAT 93950
349 349
93949 GATAAGTAAAGCTTTGGGAGAGTAGGTCAATAGCAGATTAACTGTG 93900
349 349
93899 AGTGCTTCAAAAAGACAGTTTAGAAGTAAAGTAGTAATTAGTGAGT 93850
349 349
93849 GATTAATCAGTCTGAAGTCAATAAAGAGAACTTTAATACAAATATTCC 93800
349 349
93799 GTGTAGAGAAGTCTATTAAGTTCATCATCTCAGCAGACATGGCT 93750
349 349
93749 TAAAAATTATGTTAGAGTTATTTTACAAATAAATGTACTAATGTTG 93700
349 349
93699 CTACACTAGTGAATAGTGTGTTCTATTCACCTGTTTAAAGTTAAAA 93650
349 349
93649 TAATATGATAGAGAGATAAGCTTACTGTCTCAGATGACTTAGTTT 93600
350 ArgSer11LeuSerArg13Gly 356
93599 GAAGAATATATCTTGAATTTATTTTCCAGAAATATAAACCGGCGC 93550
356 rghHisSerGluArgSerGlnArgThrGlnSerProSerLeuProAla 372
93549 GACACCACACGAAAGATCACAGAGACTCAAGGCCGCTCAGTACACAGA 93500
373 ThrProCysPheValProValProProProLeuTyProProPro 389
93499 ACTCCAGCTTTGTACTGTCTCCACACCCCTTTGTATCCGCTCTCC 93450
389 ohHisThrLeuProLeuProProGlyValProProProGlnPheSerPro 406
93449 CCATACACTTCTCTCCCTCCGGGTGTCTCTCCACAGTTTCTCTC 93400
406 lnhPheProSerSerGlnProProThrAlaGlyTyrSerValProProPro 422
93399 AGTTTCTCTCTGCGCAGCACACCCGCTGGGTATAGTGTCCCTCTCCA 93350
423 GlyPheProProAlaProAlaSerThrAlaCysPheSerProG 439
93349 GGGTTTCTCCAGCTCTGCAATTTATCAACCTTTGGGTATCAGCAG 93300
439 yValProThrAlaHisSerAsnThrMetProThrThrGlnAlaProLeu 456
93299 AGTGCAGACAGCTCATTAATACATCCCAACACACAGCACCCCTT 93250
456 euserArgGluGluPheTyrArgGluGln 465
93249 TGTCCAGGAGAAATTTATAGAGACAGCAGACTAAAGACAGTAT 93200
465 465
93199 GATTTCTAATTTGAATTTATATACACTTTTTCATTTTCTGATTAATAC 93150

```
465 ..... 465
93149 ATTAATAGTGTGGAAGCTTTCAGTGTTCAGTAGATGAGA 93100
465 ..... 465
93099 AATTCACCTAAATTTAAACCAAGCCATACAGCACTCTCAGGATT 93050
465 ..... 465
93049 CCTAGTTATAGTTTAAAGTGAACCTAATGGCAATTCATTCTTC 93000
465 ..... 465
92999 ATTGATGATCTAATTACATTTGTCTGTAGTTTCTTAATGGTTTTC 92950
465 ..... 465
92949 TTAACCTTCACCTTTTAACTTTATTTTACTAGGAAAGAAAGTCCA 92900
465 ..... 465
92899 GCTAGATGAGTTTACAAATGATTTTGTAGGAATGATGAAATACAAA 92850
466 .AsnAspLysGly..... 469
92849 AGATTCAAAAGAGCGTGGCGCTATTTCCAGGTAGTGTGTTGCAAG 92800
469 ..... 469
92799 CCTTCACACAGAAATTACAAAGGAGCTTGAATTCAGGAGATTAGCTA 92750
469 ..... 469
92749 TGGATATAAAGACATTGCTACTAGGAATTTTCTGATTCGCTAGCTG 92700
469 ..... 469
92699 TATAGAAAGTAAGATGATGATTTATTTCCCAATCTGTTCATTGAAATA 92650
469 ..... 469
92649 TGCTCAGTGTAAATTTATTTGTGTAGAACTACAGTGTGGGATTTAA 92600
469 ..... 469
92599 TTGCCCTTGCCCTTAAGTATTCATAGTATCAATATTGGGAAATGT 92550
469 ..... 469
92549 TTTTGAAAGAAAGAAATGAGGGAATACCAAAAACAGTCCCGAGACC 92500
469 ..... 469
92499 CTTTACCATAAGCTACAGCTCTCTGTGGAACATTAACTGCGTAATT 92450
469 ..... 469
92449 ATGAAGTAGTAGTGAATACAGATCGCTTTATTGATAGAAAAAGTGAATT 92400
469 ..... 469
92399 ATAAAAATTCCTGGAGTTTGTTTAAAGACTTATTTTACGCTTAC 92350
469 ..... 469
92349 ACTTGAAATGTGTTTCTTTTAAACAGCTGTGGCCAGTATAGACA 92300
469 ..... 469
92299 AAAAAGACTCTACTTCAATATTCAGATAGGAAGTGAATTTAAATTT 92250

469 ..... 469
92249 AATGACATTTTGTTAATGTAGTAGCTCATATTAACCTCAGGATTT 92200
469 ..... 469
92199 AATGCACCTCCTCAATATTGTTAAGACTTCATTTTGTGTTAAACAAGT 92150
469 ..... 469
92149 GATTGCTCCGATTTGTACTTTATTCATTCCAGTAGAGCTCAGTGC 92100
470 ..... Arg 471
92099 CTCTTAATCATCTAAGATGAATTTGATATAAGATGATGAACCGTGC 92050
471 Lu..... 471
92049 AGAAAGAAATCTAGAAATCTGAGAGTCTGTGTTTATTTGTATGTAATT 92000
472 ..... SerLysPheProty 476
91999 CTCACATATCACTAATGTTTGTGTTTGTGTTTGTAGTCTAAATCTCCCTA 91950
476 rSerGlySerSerTySerArgSerSerTyThrAspSer.SerGlnGly 492
91949 TAGTGTTCTCTGATTCAGAGAGTTCATATCTATTTCTTAATCAAGAT 91900
493 LeuAlaGlnHisIleHisAlaLeuThrLeuSerProSerAlaAlaHisIstb 509
91899 CTGGTTCACACGTTTCACGCTTATTCCTGATTCATTCACGCGCTCAGAT 91850
509 rLeuAspLeuLeuHisAspHisProHisProGluGluAlaAla 526
91849 TCTGCTCTTATTCACGCTCAGCTCCATACCCAGAAAGCAGCAGAGCAA 91800
526 rGserAlaMetIleValHisMetProAspLeuMetAspIleAlaHisAla 542
91799 GAGCCGCAATTAACGTTCCACGCTAGATCTCATGATGATATCAT. CGATCT 91751
543 ArgSerArgSerProProTyraArgTyraArgSerArgSerArgSerPr 559
91750 AGGTCAGGTCACCCCTTACAGAGCTATCATTCACGATCAAGATTCCTC 91701
559 oProGluPheArgGlyGlnSerProThrLysArgAsnValProArgGlu 576
91700 TCAAGCGTTTAGGGGACAGCTCTCTAATAAACGTAATCTACTCAAGGG 91651
576 LuLysGluArgGluTyrrPheAsnArgTyraArgGluValProProProTy 592
91650 AAACAGAACGTGAATATTTTAAATGATACAGAGAAATTCACACCAATAT 91601
593 AspIleLysAlaTyrrTyrrGlyArgSerValAspPheArgAspProPheG 609
91600 GACATGAAGCATATATATGAGGAGAGTGTGACTTTGAGAGCCCATTTGA 91551
609 uLysGluArgTyraArgGluTyrrPgluArgLysTyraArgGluTyrrPglu 626
91550 AAAAGAACGTCACCGAGATGAGAGAGAAATATAGAGAGGTATGAAA 91501
626 ystTyrrTyrrLysGlyTyraAlaValGluAlaGlnProArgProSerAlaAsn 642
91500 AATATTTTAAAGTTATGCTCTGAGACAGCTAGACCTCAGCAAT 91451
643 ArgGluAspPheSerProGluArgLeuLeuProLeuAsnIleArgAsn 659
91450 AGAGAGAACTTTCTCCAGAGAGATTTTTCGCACTTAACATCAGAAATTC 91401
659 rProPheThrArgGlyArgArgGluAspTyraAlaAlaGlyGlnSerHis 676
91400 TCCTTTCACAGAGCCGCGAGAGAACTATGTGTGGCCAAAGTCAATA 91351
676 rGAsnArgAsnLeuGlyGlyAsnTyrrProGluLysLeuSerThrArgAsp 692
```

[illegible]

```

1165 oLysgluAlaSerHisgIuleuMetGlnHisgluIleuArgSerSerLysg 1182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89512 CAAGACGTCGAGCATGAAATCATACAAATGAGGTAAAGTTCAAAA 89463
1182 LysSerAlaSerSerGluLysgluArgAlaLysAspArgGluHisSerGly 1198
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89462 ACTGTCATCTAGTGAAGAAAGGAAACCAAGATCGAGATTATTCAGTGC 89413
1199 SerGluLysAspAsnProAspLysArgLysSerGlyAlaGlnProAspLys 1215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89412 TTGGAAGAAAGAACCTGAAAGAGGAAGAACACTCGCCAGAGAA 89363
1215 sGluSerThrValAspArgLysSerGluGlnGlyHisPheSerThrLys 1232
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89362 AGAGAGTAAATTGGACCCCTCGAATGAAACAAAGAAATTTAAAGCTGCT 89313
1232 ergLysSerSerLysgluThrArgThrSerGluLysHisgluSerValArg 1248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89312 CTCATATCTTCCAAAGAGGCTAGAACGTGATTAACATGATTCACCTCGT 89263
1249 GlySerSerAsnLysAspPheThrProGlyArgAspLysLysValAspLys 1265
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89262 GCTTCTCAATTAAGACTCTCACTCCATAGAGACAAAGAAACGACTA 89213
1265 rAspSerArgAspTyrSerSerSerSerLysArgArgAspGluArgLysLul 1282
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89212 TGACACCCAGAGAGTATTCAGTTCCAAACGTAGAGATTAAGAAATGAT 89163
1282 euAlaArgArgLysAspSerProProArgGlyLysgluSerLeuSerGly 1298
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89162 TTAACAAGACGAAAGACTCTCTCTCGAATTAAGATTCTGCAATCTGGA 89113
1299 GlnLysSerLysLeuArgGluGluArgAspLeuProLysLysGly...Al 1314
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89112 CAGAAAAATTAACCAAGGGAAGAGAGATTGCTTAAAGAAACAGG 89063
1314 aGluSerLysLysSerAsnSerSerProProArgAspLysLysProHisA 1331
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89062 AGATTTCACAAAAAGTAAATTCTAGCCCTCAAGACACAGAAACCTCATG 89013
1331 sPHsLysAlaProTyrGluThrLysArgProCysGluGluThrLysPro 1347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89012 ATCAACAAAGCCACTATGATGATCAACGCCAAATGAGAGACAAATCT 88963
1348 ValAspLysAsnSerGlyLysgluArgGluLysHisAlaAlaGluAlaAr 1364
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88962 GTAGATTAAAAATCCTTGTAGAGATCGTAGAAGCATGTATTAAGACAG 88913
1364 gAsnGlyLysgluSerSerGlyAlaAsnCysHisValTyrLeuThrArg 1380
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88912 GAACAATTAAGATCAAGTGGCAATTAACCTATTATTAATTAACCCAC 88863
1381 GlnThrLeuProTyrArgArgSerTyrPleuLeuGlyArgTyrPArgArgAl 1397
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88862 CAGAGACACAGAGTTGAAAGAGCAAAATTAATGAGCAAAATGACAAAG 88813
1397 aProSerSerArgAsnProSer 1404
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88812 ACTGTCAAGCTTAAACCCAGT 88791
seq_name: gb_pr:HSRBQ1
seq_documentation_block:
LOCUS HSRBQ1 3011 bp mRNA linear PRI 29-MAY-1996
DEFINITION H.sapiens RBQ-1 mRNA.
ACCESSION X85133
VERSION X85133.1 GI:728590
KEYWORDS RB protein binding protein; RBQ-1 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

```

REFERENCE 1 (bases 1 to 3011)
AUTHORS Sakai,Y.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-1995) Y. Sakai, Biology Div., National Cancer
Center Research Inst., Tsukiji 5-1-1, Chuo-ku, Tokyo 104, JAPAN
REFERENCE 2 (bases 1 to 3011)
AUTHORS Sakai,Y., Saijo,M., Coelho,K., Kishino,T., Nlkawa,N. and Taya,Y.
TITLE cDNA sequence and chromosomal localization of a novel human
protein, RBQ-1 (RBBP6), that binds to the retinoblastoma gene
product
JOURNAL Genomics 30 (1), 98-101 (1995)
MEDLINE 96129310
FEATURES
source
location/Qualifiers
1..3011
/organism="Homo sapiens"
/strain="NCI-H69"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p11.2-12"
/clone="1"
/cell_line="NEC"
/cell_type="small cell lung carcinoma"
/clone_1db="lambda ZAP1"
1..2994
/gene="RBQ-1"
/product="RB protein binding protein"
/evidence="experimental"
1..2994
/gene="RBQ-1"
92..2938
/gene="RBQ-1"
/codon_start=1
/product="RB protein binding protein"
/protein_id="CAA59445.1"
/db_xref="GI:755748"
/db_xref="SPTREMBL:Q15290"
/translacion="MGIKTLNLYVLGLRALLEPEYFMEEKOPNNKGAMLTNTGKVAL
PIDAEAYVIGKKRPPPLPEPPSSSESDPIPELCLICKDINTDAVVPCCGNS
YCDCEIRATLALSDHEKTCPTCHONOVSPALANKFLROAVNNKREKGYTRLRKLQ
PPPPPIPPRPLIORNIQPLMRPSIRRODPLMIIPVSS7HPAPSSISLNSQSL
APPVSGNPSAPAPPDITATYVTSVSHSKSGPPRDSNKLTPAALASBSKSTSS
IAITLAMEKGYOVPLGTPSLGSLGSLHGLIPTYGPRVINTARGGGRPCEWSNK
LGVLVSPPOOIRGERSQCYRSINRGHNSERQOGPSLPATPVFVPPVPPPLVPP
PHTLPPVPPPOEFPPOCPGPPAGSVPPGFPAPAPAMLSIPVWSSGVOTASHN
TIPYQAPPLSREEFYREORLKEEKSKSLDEFTNDPAKELMEYKIOKRRRSFS
BKSPVSGSSYSRSYVTSYRSRSRSGSTRSGRSFSRSRSYSPVPRRGKSR
NYSRSRSGHYHRSRSRSPYRHYHRSRSPAPRQSPNNKNNVQGETREYENRNR
EVPPEYDMKAVYGRSVDFRDEKERTREWERKRWETKTKYKGTAGAGPRPSARE
NESPENFLPLNIRNSPFTGRREDYVGGSHSRNIGSNYPKLSARDGHNOKDKWS
KEKESNADGDKGNKHKHRRKRKESEEGFLNPELLETSSRSKSPTEGVEBKDLSL
FVLPSHDDATPVDEPMDAESTIFTSVSEKDRERDKPRAKGDYTRKKMDGSAVSKK
NIVKPAKGPOEKVDGVDRLDLNLOLTKPKRRRLRLTLNHLPLRRKKKSLPEPEK
LTLNODKTRNKTSQDKSEBGLFQRCQIRKANN"
BASE COUNT 1034 a 682 c 600 g 695 t
ORIGIN
alignment_scores:
Quality: 3938.00 Length: 978
Ratio: 4.633 Gaps: 8
Percent Similarity: 86.912 Percent Identity: 78.937
alignment_block:
US-09-811-045A-1 x HSRBQ1 ..
Align seg 1/1 to: HSRBQ1 from: 1 to: 3011
1 MetwctGluValLysAspProAsnMetLysGlyAlaMetLeuThrAsnTh 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
158 ATGATGAGAGTGAAGATCCCTATATGAAAGGTCAATGCTTACCACAC 207
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 rGlyLysTyrAlaIleProThrIleAspAlaGluAlaTyrAlaIleGlyL 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

208 TGGAAATATGCAATACCACTATAGATGACAGACATATGCAATTGGGA 257
34 yLysGluYrsProProPheLeuProGluGluProSerSerSerGlu 50
258 AGAAAGAAACCTCCCTCTTACAGAGAGACCATCTTCTCCACAGA 307
51 GluAspAspProIleProIaGluLeuLeuCysLeuIleCysLysAspIle 67
308 GAAGATGATCCATCCAGATGATGATGTTGTCTCATCTGCAAGATAT 357
67 eMetHrAspAlaValIleProCysCysGlyAsnSerSerCysAspG 84
358 TATGATGATGCTGTGTGATTCCTCGTGTGGAACAGTACTGTGATG 407
84 LucYstIleArgHrThrLeuLeuGluSerAspLysHisThrCysProHr 100
408 AATGATTAAGAACAGCATCTGGAATCAGATGACACACATGTCGACG 457
101 CysHisGlnAsnAspValSerProAspAlaLeuIleAlaAsnLysPhe 117
458 TGTGATCAAAATGATGTTCTCTGATGCTTAAATGCCAATAATTTT 507
117 uArgGlnAlaValAsnAsnPhelYsnGluThrGlyTyrHrLysArgL 134
508 ACGACAGCTGTAATAACTCAAAATGAAACTGSGCTATACAAAAGAC 557
134 euArgGlySlnLeuProProPheLeuPheLeuValProProProArgPro 150
558 TACGAAAGAGTTACTCTCCACACACCCCAATACCACTCCGAGACCA 607
151 LeuSerGlnArgAsnLeuGlnProArgSerArgSerProIleLeuArgL 167
608 CTGATTCAGAGGACCTACACCTCGATGAGATCTCCGATATACAGACA 657
167 nGlnAspProValValPheArgTyrThrValSerProHrCysSerAspT 184
658 ACAAGATCCTCTTATGATGACAGTGCATCTTCATCACT..... 697
184 hrLysHrIaGlySerCysSerAspSerGlyThrLeuSerArgLeuPro 200
698CACCA 703
201 AlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAlaProYr 217
704 GCTCGGTATATCTTCATTACTCTTAATCAAGTCTCTGCCCCCTCC 753
217 oValSerGlyAsnProSerSerAlaProAlaProValProAspIleHrTha 234
754 TGTGTCTGAAATCCGCTTCTGCTCCACCTCTGATCTGATATACAG 803
234 IaThrValSerIleSerValHisSerGluLysSerAspGlyProPheArg 250
804 CACAGATATCCATATCAGTTTCATTCAGAAATCAGATGACCTTTCCG 853
251 AspSerAspAsnLysLeuLeuProAlaAlaAlaLeuThrSerGluHisSe 267
854 GATTCGATAATAATATGTCAGCTGCAGCTGCACTTCGACAGAGACAC 903
267 rLysGlyAlaSerSerIleAlaIleHrAlaLeuMetGluGluLysGlyY 284
904 AAAGGGAACCTCTCAATGCAATACCCCTCTTATGGAAGAAAGGTTT 952
284 alProGlyHrSerProTrpAsnSerIlePheValGlyGlnSerLeuLe 300
953 A.CCAGGTGCTGTTCTTGGAAACCCCATCTTTCCTTGACAGTCATTA 1002
300 uHisGlyGlnLeuIleProHrThrGlyProValArgIleAsnAlaAla 317
1003 GCATGACAGTGTGATCCACAACTGTCGACGTAAGATAATATGTCGC 1052
317 rGProGlyLysGlyArgProGlyTyrGluHisSerAsnLysLeuGlyTyr 333
1053 GTCCAGGTGTGTGTCAGCAGCTGGGAACATTCAACAACTTGGCTAT 1102

334 LeuValSerProProGlnGlnIleArgArgGlyGluArgSerCysTyrAr 350
1103 CTGCTTCTCCACCACACAAATTAAGAAAGGAGAGCTGCTACAG 1152
350 gSerIleAsnArgGlyArgHisHisSerGluArgSerGlnArgThrGlns 367
1153 AAGTATAAACCGTGGGCGACACACAGCGAAGATCACAGAGACCTCAAG 1202
367 erProSerLeuProAlaHrProCysPheValProValProProProPro 383
1203 GCCCGTACCTACAGCACTCCAGTCTTGTACTGTCCTCCACACCTCTCT 1252
384 LeuTyrProProProHrHisThrLeuProLeuProGlyValProPr 400
1253 TTGTATCCGCTCTCCATACACTTCTCTCCCTCCGGGTCTCTCC 1302
400 oProGlnPheSerProGlnPheProSerSerGlnProProHrAlaGlyT 417
1303 TCCACAGTTTTCCTCCACATTTCTCTGCGCACACACACCGCTGGGT 1352
417 ySerValProProProGlyPheProProAlaProAlaAsnIleSerThr 433
1353 ATAGTCTCCCTCTCCAGGTTTCTCCAGCTCTCCCAATTTATACACA 1402
434 AlaCysPheSerProGlyValProThrAlaHisSerAsnThrMetProH 450
1403 CCTGGGTATCATCAGAGAGTGACAGACGCTCATCAAAATCAATCCAC 1452
450 rThrGlnAlaProLeuLeuSerArgGluPheTyrArgGluGln..... 465
1453 AACACAGACACACCTTGTGCGAGAGAAATCTTATAGACACACGAC 1502
466Asn 466
1503 GACTAAAGAGAGAAAGAAAGAAAGTCCAGCTAGATGATTACAAAT 1552
467 AspLysGlyArgGlu..... 471
1553 GATTTGCTAAGAAATTGATGAAATCAAAAGATTCAAAAGAGCGCTAG 1602
472SerLysPheProTyrSerGlySerTyrSerA 483
1603 GCGCTCATTTTCCAGGTCTAAATCTCCCTATAGTGTCTGCTATTCAA 1652
483 rGSerSerTyrThrAspSerSerGlnGlyLeuAlaGlnHisIleHisAl 499
1653 GAAGTTCATATACTTATTTCTAAATCAAGATCTGCGTCAACACGTTCA 1702
499 alLeuThrLeuSerProSerAlaAlaHisThrLeuAspLeuLeuAspH 516
1703 TCTTATTCGATCATTCACGCGCTCACATTCCTGCTTCATTCAGGTC 1752
516 IaProHisProProGluGluAlaGluAlaArgSerAlaMetIleValHis 532
1753 ACCCTCATACCCCAAGAGAGGAGGAGCAAGACCCCAATTACCGTTC 1802
533 MetProAspLeuMetAspIleAlaHisAlaArgSerArgSerProPro 549
1803 GGTCTGATCTCATGATATCAT.CGATCTAGGTCAAGGTCACCCCTTA 1851
549 rArgArgTyrArgSerArgSerArgSerProProGluPheArgGlyGlns 566
1852 CACAGCGCTATCATTCACGATCAAGATCTCTCAAGGCTTTAAGGGACAG 1901
566 erProThrLysArgAsnValProArgGluGluLysGluArgGlyTyrPhe 582
1902 CTCTAATAAAGCTAATGTACTCAAGGGAACAACAGCTGAATATTTT 1951
583 AsnArgTyrArgGluValProProProTyrAspIleLysAlaTyrTyrG 599
1952 AATGATATACAGAGAGTTCACACCATATGACATGAAGCAATATTTAG 2001

82 yAspGluCysIleAArgThrThrLeuLeuGluSerAspLysHisThrCys 98
 |||||.....
 8491 GTCATGAAATGTTAAAGACAGCACTTCTGAGTCAGATGACATACATCATGT 8442
 99 ProThrCysHisGlnAsnAspValSerProAspAlaLeuIleAlaAsnLys 115
 |||||.....
 8441 CCAACATGTCATCAAAATGATGTTCTCTGATGCTTAGTTAGTCCAAACA 8392
 115 sPheLeuArgGlnAlaValAsnAsnPhelysAsnGluThrGlyTyrThL 132
 |||||.....
 8391 GGTTTTACAGACAGGCTGTTAAATTAATTTAAATCAAACTGACTACACAA 8342
 132 yAArgLeuArgGlyGlnLeuProProPheLeuPheLeuValProProPro 148
 |||||.....
 8341 AGAGACTGGGAAAACAGGCTACTCCGTCGCCACCCCACTTACTTCCACCA 8292
 149 ArgProLeuSerGlnArgAsnLeuGlnProArgSerArgSerProIleLe 165
 |||||.....
 8291 ACTGCATCATTCAGACAGAACCTGCAGCCTCCTATGAAATCTCTCAGATC 8242
 165 uArgGlnGlnAspProValValPheArgTyrThrValSerProThrCys 182
 |||||.....
 8241 TAGACAGCAGGATCCTTGAAAGATTCCAATGACA..... 8208
 182 eAspThrLysThrAlaGlySerCysSerAspSerGlyThrLeuSerArg 198
 |||||.....TCGCTCAGCCTCAGCT 8190
 8207TCGCTCAGCCTCAGCT 8190
 199 LeuProAlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAl 215
 |||||.....
 8189 CACCAACCTCCTGTGTAACCTCATAGCTCAAAATTCATCTTCCTCGGC 8140
 215 aProProValSerGlyAsnProSerSerAlaProAlaProValProAspI 232
 |||||.....
 8139 TCCCTCTGCTGGAAACCACTCTCTCCAGCTCCGTAACCGGACA 8090
 232 LeThrAlaThrValSerIleSerValHisSerGluLysSerAspGlyPro 248
 |||||.....
 8089 CCACCTGCAGAGATCTGTATCAGACCTTCAGAAAATCAGATGATGC 8040
 249 PheArgAspSerAspAsnLysLeuLeuProAlaAlaAlaLeuThrSerG 265
 |||||.....
 8039 TTTCCGGAATCAGAAAACCAATATATACAGCGCCGCCCTTACATCAGA 7990
 265 uHisSerLysGlyAlaSerSerIleAlaIleThrAlaLeuMetGluL 282
 |||||.....
 7989 ACATTCAAAGGAGCCTCTTCATTCGTGTACTGCTCTATGGAAGAAA 7940
 282 ySgLYValProGlyThrSerProThrAsnSerIlePheVal_GlyGlnSe 298
 |||||.....
 7939 ACGCTGG.CCAGGTGGCGTCTCTGAACCTCCACCTTCGTTGGGACAGTC 7891
 298 rLeuLeuHisGlyGlnLeuIleProThrThrGlyProValArgIleAsn 315
 |||||.....
 7890 ACTACTCCATGAACAGTCTATCCCCACAACCTGTCAGTAAGAAATAAG 7841
 315 lAlaLArgProGlyGlyArgProGlyTyrGluHisSerAsnLysLeu 331
 |||||.....
 7840 CTGCTCATCCAAAGTGTGTCAACAGGTTGGGAACTTCCAACAAGATT 7791
 332 GlyTyrLeuValSerProGlnGlnIleArgArgGlyGlyLysSerG 348
 |||||.....
 7790 GGGAGCTAGTTTCTTCACACAGCAAAATTAGAGTGAAGAAAGAGCTG 7741
 348 sTyrArgSerIleAsnArgGlyLysHisSerGluArgSerGlnArgT 365
 |||||.....
 7740 TTACCGAGAAATPAAACGTGGGAGACTAGTAAGTAAACATCACAGAG 7691
 365 hGlnIleProSerLeuProAlaThrProCysPheValProValProPro 381
 |||||.....
 7690 GTCAAGGCCATCAGACAGCAAGCTCAGTCTTGTGCTATTCACACA 7641
 382 ProProLeuTyrProProProProHisThrLeuProLeuProProGlyVal 398

|||||.....
 7640 CCTCTTTCCACCGCTCTCCACATATGCTTCCCTCCACAGGTGT 7591
 398 lProProProGlnPheSerProGlnPheProSerSerGlnProProThr 415
 |||||.....
 7590 ACTGCCCCACGGTTTCCCTCCGGGTTCTCTCCGCGCCAGCTCCAAACAG 7541
 415 lAgLYTyrSerValProProProGlyPheProProAlaProAlaAsnIle 431
 |||||.....
 7540 CAGGATATAGTCTCCCTCCAGGATTTCCACACAGCTCTCCCAATATA 7491
 432 SerThrAlaCysPheSerProGlyValProThrAlaHisSerAsnThr 448
 |||||.....
 7490 TCACAGCTTGCTTTTACACAGAGGTGTCCACAGCCCAATCAATACAT 7441
 448 tProThrThrGlnAlaProLeuLeuSerArgGluGluPheTyrArgGlu 465
 |||||.....
 7440 GCCCAACACAAAGCAGCTTTTGTCCAGGAGAAATTTATAGAGAGC 7391
 465 lAsnAspLysGlyArgGluSerLysPheProTyrSerGlySerSerTyr 481
 |||||.....
 7390 AAAACGACTAAAGAGAGTCTAAATTTCCCTATAGTGGTCATCGAT 7341
 482 SerArgSerSerTyrThrAspSer_SerGlnGlyLeuAlaGlnHisIleH 498
 |||||.....
 7340 TCAGAGAGTTCAATACACCTGACTCAAGTCAAGGTGGCTCAGACATTC 7291
 498 lAlaLeuThrLeuSerProSerAlaAlaHisThrLeuAspLeuHis 514
 |||||.....
 7290 ACGCTCTTACTCAGATCCTTCAAGCTGCTCACACCTCTCATCTTCTCAC 7241
 515 AspHisProHisProProGluGluAlaGluAlaArgSerAlaMetIleVal 531
 |||||.....
 7240 GATCATCCCATCTCTCCAGAGGCGAGGCAAGATCTGCAATGATGT 7191
 531 lHisMetProAspLeuMetAspIleAlaHisAlaArgSerArgSerPro 548
 |||||.....
 7190 TCACAGCTCAGATCTCATGATATGCCCATCTC_AAGTCAAGGTCCCCC 7142
 548 tGlyTyrArgGlyTyrArgSerArgSerArgSerProProGluPheArgGly 564
 |||||.....
 7141 CCTATGAGAGATATCTCAGCGTCCAGATCTCTCCACATTTAGGGGA 7092
 565 GlnSerProThrLysArgAsnValProArgGluGluLysGluArgGly 581
 |||||.....
 7091 CAGTCTCCCACTAACTCGATGACTCTCAGAA..... 7059
 581 rPheAsnArgTyrArgGluValProProProTyrAspIleLysAlaTyrT 598
 |||||.....
 7059 7059
 598 ySgLYArgSerValAspPheArgAspProPheGluLysGluArgTyrArg 614
 |||||.....
 7058GAGAAAGACCTTACCAG 7041
 615 GluThrGluArgLysTyrArgGluTyrGlyLysTyrTyr_LysGlyTyr 631
 |||||.....
 7040 GAATGGAGAGAAATATCACTAGTGTATGAGAAAT...AGGCCAAAGGCC 6994
 6993 AAGCATTAGGAACACAGCGCTAGACCTTAGGCCAATAGAGAGGACTTTCT 6944
 648 ProGluArgLeuLeuProLeuAsnIleValGlnSerSerProPheThrArg 664
 |||||.....
 6943 TCAGAGAGACTTACTCTCTGATATCAGAAATTTCAACCTTCAAGAGAG 6894
 664 yArgArgGluAspTyrAlaAlaGlyGlnSerHisArgAsnArgAsnLeu 681
 |||||.....
 6893 GCACAGAGAGACTATGCTGTGACAAAGTCAATAGAAATAGAAAGCTTGA 6844
 681 LysGlyAsnTyrProGluLysLeuSerThrArgAspSerHisAsnAlaLys 697
 |||||.....

```

6843 GTGGGACTGTCACGTAAGCTTTCAACAGACAGACATCCAAATCCAAA 6794
698 AAGPAPProLysSerLysGluSerGluAsnValProGlyAspG1 714
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6793 GATTAATCCAAATTCGAAAGAGAAAGAGACATGCTTCTCGAGACAG 6744
714 yLysGlyAsnLysSHS 719
:|||||:|||||:
6743 CAAGAGAACACAGAAAT 6728

seq_name: gb_hhg:AC099701
seq_documentation_block:
LOCUS AC099701 222549 bp DNA linear HTG 18-NOV-2001
DEFINITION Mus musculus clone RP23-184F6, WORKING DRAFT SEQUENCE, 21 unordered
pieces.
AC099701.1 GI:16974200
VERSION AC099701.1
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus.
1 (bases 1 to 222549)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-184F6
Unpublished
2 (bases 1 to 222549)
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Chopel,Y., Colangelo,M., Collins,S., Collingore,A., Cook,A.,
Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gargana,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Laroque,K.,
Lamazeas,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,
Menous,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Rella,R., Rieback,M., Riley,R., Risse,C., Rogov,P.,
Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schnapack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggilo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (18-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L17215
Center clone name: 184_F-6

----- Summary Statistics
Sequencing vector: Plasmid; n/a: 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 213468 bases at least Q40
Consensus quality: 218282 bases at least Q30
Consensus quality: 219801 bases at least Q20
Insert size: 220549; sum-of-contigs
Quality coverage: 5.8 in Q20 bases; sum-of-contigs

```

```

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
861: contig of 861 bp in length
862 961: gap of 100 bp
962 1566: contig of 605 bp in length
1567 1666: gap of 100 bp
1667 2720: contig of 1054 bp in length
2721 2820: gap of 100 bp
2821 4833: contig of 2013 bp in length
4834 4933: gap of 100 bp
4934 7443: contig of 2510 bp in length
7444 7543: gap of 100 bp
7544 10612: contig of 3069 bp in length
10613 10712: gap of 100 bp
10713 13264: contig of 2552 bp in length
13265 13364: gap of 100 bp
13365 15293: contig of 1929 bp in length
15294 15393: gap of 100 bp
15394 20002: contig of 4609 bp in length
20003 20102: gap of 100 bp
20103 26889: contig of 6787 bp in length
26890 26989: gap of 100 bp
26990 33039: contig of 6050 bp in length
33040 33139: gap of 100 bp
33140 38303: contig of 5164 bp in length
38304 38403: gap of 100 bp
38404 46409: contig of 8006 bp in length
46410 46509: gap of 100 bp
46510 54712: contig of 8203 bp in length
54713 54812: gap of 100 bp
54813 70402: contig of 15590 bp in length
70403 70502: gap of 100 bp
70503 86873: contig of 16371 bp in length
86874 86973: gap of 100 bp
86974 101977: contig of 15004 bp in length
101978 102077: gap of 100 bp
102078 128046: contig of 25969 bp in length
128047 128146: gap of 100 bp
128147 161179: contig of 33033 bp in length
161180 161279: gap of 100 bp
161280 188228: contig of 26849 bp in length
188229 188328: gap of 100 bp
188329 222549: contig of 34221 bp in length.
Location/Qualifiers
1..222549
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-184F6"
/clone_lib="RP23-184F6"
1..861
/note="assembly_fragment"
vector_side:left"
clone_end:SP6
662..1566
/note="assembly_fragment"
1667..2720
/note="assembly_fragment"
2821..4833
/note="assembly_fragment"
4934..7443
/note="assembly_fragment"
7544..10612
/note="assembly_fragment"
10713..13364
/note="assembly_fragment"
13365..15293

```

```
misc_feature /note="assembly_fragment"
15394..20002 /note="assembly_fragment"
misc_feature /note="assembly_fragment"
20103..26889 /note="assembly_fragment"
misc_feature /note="assembly_fragment"
26990..33039 /note="assembly_fragment"
misc_feature /note="assembly_fragment"
33140..36303 /note="assembly_fragment"
misc_feature /note="assembly_fragment"
38404..46409 /note="assembly_fragment"
misc_feature /note="assembly_fragment"
46510..54712 /note="assembly_fragment"
misc_feature /note="assembly_fragment"
54813..70402 /note="assembly_fragment"
misc_feature /note="assembly_fragment"
70503..86673 /note="assembly_fragment"
misc_feature /note="assembly_fragment"
86974..101977 /note="assembly_fragment"
misc_feature /note="assembly_fragment"
102078..128046 /note="assembly_fragment"
misc_feature /note="assembly_fragment"
128147..161179 /note="assembly_fragment"
misc_feature /note="assembly_fragment"
161280..188228 /note="assembly_fragment"
misc_feature /note="assembly_fragment"
188329..222549 /note="assembly_fragment"
BASE COUNT 63391 a 45670 c 46003 g 65450 t 2035 others
ORIGIN
```

```
alignment_scores:
Quality: 2576.50 Length: 725
Ratio: 4.404 Gaps: 9
Percent Similarity: 80.690 Percent Identity: 72.414
```

alignment block:
US-09-811-045a-1 x AC099701 ..

Align seg 1/1 to: AC099701 from: 1 to: 222549

```
1 MetMetGluValIlyAspProAsnMetIySgLYAlaMetLeuThrAsn 17
|||||
148226 ATGATGGAAGTGAAGATCCTATACATGAAGAGTGCAATGCTTCAAAAAC 148275
|||||
17 rGlyIySTYrAlaIleProThrIleAspAlaGluAlaTYrAlaIleGlyL 34
|||||
148276 TGGGCAATATGCAATACCGACTTAAATGACAGAGGCTATGCAATTGGGA 148325
|||||
34 ySlySgLIuIySProPhoPheLeuProGIuIuPro.....SerSerSer 48
|||||
148326 AGAAAGAGAAACACCCCTTCTTACAGAGAGAACCATCATCATCATCATCT 148375
|||||
49 SerGIuGIaSPasPProIleProAlaGluLeuLeuCysLeuIleCysLy 65
|||||
148376 TCAGAGAGAGTTCATCTGTCGCCAGACGAGCTTGTGCTCATCTGCGCA 148425
|||||
65 sasPIleMetThrAspAlaValIleProCysGlyAsnSerSerC 82
|||||
148426 AGACACTATGACTGATGCTGCTATCATCCCTGCTGGAACAGATTACT 148475
|||||
82 yAspSgLIuIySgLIeArgThrThrLeuLeuGluSerAspLySHstHrCys 98
|||||
148476 GTGATGAATGTATAAGAACGACCTTCTGGAGTCAGATGAACATACATCT 148525
|||||
99 ProThrCysHisGlnAsnAspValSerProAspAlaLeuIleAlaAsnLy 115
|||||
148526 CCAACATGTCATCAAAATGATGTTTCTCCGATGCTTAACTGCGCAACA 148575
|||||
115 sPheLeuArgGlnAlaValaIaAsnAsnPhelySasngIuThrGlyTYrThrL 132
|||||
148576 GGTTTTACGACAGCGCTGTTAAATACTTTAAATAACAACGACTGCACAA 148625
|||||
```

```
132 yArGLeuArgIySgLIuLeuProPhoPheLeuPheLeuValProPro 148
|||||
148626 AGAGGCTCGGAAAACAGGTCACCTCCCTCCCAACCCCATCTACTACACA 148675
|||||
149 ArgProLeuSerGIaArgAsnLeuGlnProArgSerArgSerProIleLe 165
|||||
148676 AGTGCATCATTCAGCAGAACCTGCAGCTCCATGAATCTCCACATC 148725
|||||
165 uArgGInGlnAsPProValaIaPheArgTYrThrValSerProThrCys 182
|||||
148726 AAGACAGCAGGATCCTCGAAGGTTCCAGTGACA..... 148759
|||||
182 eArSPHrIySTHrAlaGlySerCysSerAspSerGlyThrLeuSerArg 198
|||||
148760 .....TCGTTCTCAAGCT 148771
|||||
199 LeuProAlaProSerIleSerIleSerLeuThrSerAsnGlnSerSerLeu 215
|||||
148772 CACCCAACTCCCTCTGTAACCTCATAGCTTCMAATTTATCTTCCTCGCG 148821
|||||
215 aProProValSerGIyAsnProSerSerAlaProAlaProValProAsp 232
|||||
148822 TCCCTCTGCTCGAACCACCATCTTCTGCCAGCTCCAGTACGTAGATA 148871
|||||
232 IeThrAlaThrValSerIleSerValHisSerGluIySerAspGlyPro 248
|||||
148872 CAACCTCAGAGATATCTATATTCAGACCATTCAGAAAATCAGATGCARCC 148921
|||||
249 PheArgAspSerAspAsnIySgLIuLeuProAlaAlaIleuThrSerGI 265
|||||
148922 TTTCGGGATTCAGATTAACAATATATATTCAGCTGCCGCTTATCATCAGA 148971
|||||
265 uHisSerIySgLYAlaSerSerIleAlaIleThrAlaLeuMetGluIu 282
|||||
148972 ACAATTAAAGGAAGCCCTTCAATTCGTGCTGCTTATTAACAAGAAA 149021
|||||
282 ySgLYAlaProGIyThrSerProTrpAsnSerIlePheValGly..Glnse 298
|||||
149022 AGTGCTAG..CCAGGTGCCAGTCTTGAACCTTCACCTTTGTTGGACAGATC 149070
|||||
298 rLeuLeuHisGlyGlnLeuIleProThrThrGlyProValArgIleAsn 315
|||||
149071 ACTATTACATGAACAGATTATATCCCACTGCTCATTAAGAATAATG 149120
|||||
315 IaAlaArgProGIyGlyArgProGIyTrpGluHisSerAsnIySleu 331
|||||
149121 CTGCTCATTCAAATGCTGTGTCACAGAGTGGAGACATCCCAACAAGCTT 149170
|||||
332 GlyTYrLeuValSerProProGInGlnIleArgArgGlyGluArgSerC 348
|||||
149171 GGGAGCTAGTTCCTCCACACAGCAAAATTAAGAGTGAAGAAAGACTCTG 149220
|||||
348 sTYrArgSerIleAsnArgGlyArgHisHisSerGIuArgSerGlnArg 365
|||||
149221 TTACCGAGGAATAAAGTGGGCAAGACTACAGTGAACATCACAGAGCA 149270
|||||
365 hGlnSerProSerLeuProAlaThrProCysPheValProValProPro 381
|||||
149271 GTCAAGGCCACACGACGACGACCACTCCATCTTGTGCTATTCACCA 149320
|||||
382 ProProLeuTyPProProProHisThrLeuProLeuProProGIyVal 398
|||||
149321 CCTCTTATGCAACTGCTCTCCACATATGCTTCTTCTCTCCAGGTGT 149370
|||||
398 lProProProGlnPheSerProGlnPheProSerSerGlnProProTha 415
|||||
149371 ACCTGCCGCCAGGTTCCCTCTGTTTCTCTCTAGCAGACCTCCACAG 149420
|||||
415 IaGlyTYrSerValProProProGIyPheProProAlaProAlaAsnIle 431
|||||
149421 CAGGATATATGTCCTCCCTCCAGAGATTCACACACTCTCTGCAATATA 149470
|||||
432 SerThrAlaCysPheSerProGIyValProThrAlaHisSerAsnThrme 448
|||||
```

```

||||||| ||| ||||| |||||||||||||||||
149471 TCACACCTGGTGGTTTCATGCAGAGTTCACACTGCTCATTCATAATACCAT 149520
448 tProthThGlnAlaProLeuLeuSerArgGluGluPheYrArgGluG 465
:|||||||||||||||||||||||||||||||||
149521 CCCCAACAACACACACCTTTGTCAGGGAAGAATTCATATGAGAGC 149570
465 InaNaApLySGlyArgGluSerLySPheProYrSerGlySerSerYr 481
||||| :|||||
149571 AAAACACGATRAAGAAAGAGCTAAATTCCTTATAGTGGTCAATCGAT 149620
482 SerArgSerSerYrThrAspSer.SerGlnGlyLeuAlaGlnHisLeH 498
||||| |||||||
149621 TCAGAAGATTCATACACTGACTCAAGTCAGGCTGCTCAACACATTC 149670
498 lSaIaLeuThLeuSerProSerAlaAlaHisThrLeuAspLeuHis 514
||||| |||||||
149671 AGGCTCTTACTCTAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 149720
515 AspHisProHisProProGluGluAlaGlnAlaArgSerAlaMetIleVa 531
||||| |||||||
149721 GATCATCCCATCTCCAGAGAGGCAAGCATCTGCATGATGATGT 149770
531 lHisMetProAspLeuMetAspIleAlaHisAlaArgSerArgSerProp 548
||||| |||||||
149771 TCACAGTCCAGATCTCATGTGATATCAC.CGATCTAGGTCAGAGTCCCTC 149819
548 roYrYrArgYrArgSerArgSerArgSerPropGluPheArgGly 564
||||| |||||||
149820 CCTATAGACGATATACACTCACGCTCCAGATCTCTCCAGCTTTAGGGGA 149869
565 GlnSerProThrLyArgAsnValProArgGluGluLySGluArgGluTy 581
||||| :|||||
149870 CAGCTCCCATTAACCTGATCTGACTCAAGAA..... 149902
581 rPheAsnArgTYrArgGluValProProProTYrAspIleYsAlaTYrT 598
149902 ..... 149902
598 YrGlyArgSerValAspPheArgAspProPheGluLySGluArgTYrArg 614
149903 .....GAGAAAGAACCTACACAG 149920
615 GluTrpGluArgLyTYrArgGluTrpTYrGluYsTYrTYr.LysGluTY 631
149921 GAATGGGAGAGGAATACTACACAGGTATGAGAT...AGCAAAAGGCGC 149967
631 YrAlaValGlyAlaGlnProArgProSerAlaAsnArgGluAspPhe.Se 647
149968 AAGCAGATAGGAACACAGCTAGACCTTAGGCCAATAGAGAGTTTCTTC 150017
647 rProGluArgLeuLeuProLeuAsnIleArgAsnSerProPheThrArgG 664
150018 A...GAGAGACTTTCCTCTTAATATCAGAAATTCATCTTCACAGAG 150064
664 LyArgArgGluAspTYrAlaAlaGlnSerHisArgAsnArgAsnLeu 680
150065 GGCACAGAGAGAGACTATGCTGTCGACAAAGTCATAGAAATAGAAATCTA 150114
681 GlyGlyAsnTYrProGluLyLeuSerThrArgAspSerHisAsnAlaLy 697
150115 GGTGGCACTTCACAGTAAAGCTTTCAACAAGAGACAGTCACAATGCAAA 150164
697 sAspAsnProLySerLySGluYsGluSerGluAsnValProGlyAspG 714
150165 AGATATTCCTCAAAATCGAAAGAGAGAGATCAGAAATGTTCTTCGACAGCA 150214
714 lYLySGlyAsnLySHis 719
:|||||
150215 GCAAAAGAAACAGAGAT 150231
seq_name: gb_pat:AR048215

```

```

seq_documentation_block:
LOCUS AR048215 1800 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 11 from patent US 5821070.
ACCESSION AR048215
VERSION AR048215.1 GI:5970558
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1800)
AUTHORS Lee, W.-H. and Shan, B.
TITLE Antibodies reactive with retinoblastoma binding proteins and
METHODS of using same
JOURNAL Patent: US 5821070-A 11 13-Oct-1998;
FEATURES Location/Qualifiers
source 1..1800
BASE COUNT 747 a 333 c 372 g 348 t
ORIGIN

alignment_scores:
Quality: 1838.50 Length: 490
Ratio: 4.113 Gaps: 5
Percent Similarity: 91.224 Percent Identity: 75.510

alignment_block:
US-09-811-045A-1 x AR048215 ..
Align seg 1/1 to: AR048215 from: 1 to: 1800

919 ProArgSerLeu...ArgLySIlEaSnTYrLeuIlEaAlaArgGluLySAs 934
11 CCAAGAGCCCTAATGAGAAAAAACAACACACTTGATATATAGGAGGAAAAA 60
934 nGUaRGluLySArgLyS..LySerValAspLySAspPheGluSerSe 950
:||||| ||| ||| :|||||
61 AGAAAGAAAGAACTGAAAGAAAGGCGTGAATAGTAAAGTTTGACTTTC 110
950 rSerMetLySIlESeLySValGluGlyThrGluIlEValLySProSerP 967
111 TTCATATGAATCTCGAAACTAGAGAGTACGATGATATAGTAAACATCAC 160
967 roLySArgLySMetGluGlyAspValGluLySLeuGluArgTYrProGlu 983
161 CAAAGCGCAAAATGGAACCTGATCTGAAAAAATGATAGACCCCTGAA 210
984 LySAspLySIlEaLaserSerThrThrProAlaLySIlElySLeuAs 1000
||||| ||| :|||||
211 AAGACCAAAAT...TCTTAAAGTGGCGCAGCCCAAAAATCAAACTCA 257
1000 nArgGluThrGlyLySIlEgLyAsnAlaGluAsnAlaSerThrTHT 1017
258 CAGAGAAACTGGAAGAAATTTGGAAGTACAGAAATATATCAAAACAA 307
1017 ySGluProSerGluLySLeuGluSerThrSerSerLySIlElySGlnGlu 1033
308 AAGAACCCTCGAAAAATTTGGAGTCAACATCTACCAAAATTAACAAGAA 357
1034 LySValLySGlyLySAlaLySArgLySValAlaGlySerGluGlySerSe 1050
358 AAAGTCAAAGGAAAGGTCAGACGAAAGTGAAGTGAAGTGAAGATCTAG 407
1050 rSerThrLeuValAspTYrThrSerThrSerSerThrGlyLySerProV 1067
408 CTCACCTCTGGTGAATTCACACAGTACAGCTCAACTGAGGAGCTCTG 457
1067 aLArgLySserGluGluLySThrAspThrLySArgTYrValIlElySThr 1083
458 TGGGAAATCTGAAGAAAAAAGACATCAAGGCAAGCTGATTAAGAAC 507
1084 MetGluGluTYrAsnAsnAspAsnThrAlaProAlaGluAspValIlEIl 1100
|||||

```



```
----- Project Information
Center project name: GCQM
Center clone name: CH230-8F21
----- Summary Statistics
Assembly program: Phrap; version 0.990329Ffirst call to
findPhrapList
Consensus quality: 121430 bases at least Q40
Consensus quality: 129663 bases at least Q30
Consensus quality: 136708 bases at least Q20
Estimated insert size: 119515; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1.6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank-draft-data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 68 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
6858: contig of 6858 bp in length
6859: gap of unknown length
6959: contig of 5759 bp in length
12717: contig of 5759 bp in length
12817: gap of unknown length
16979: contig of 4162 bp in length
16980: 17079: gap of unknown length
17080: 21262: contig of 4183 bp in length
21263: 21362: gap of unknown length
21363: 25724: contig of 4362 bp in length
25725: 25824: gap of unknown length
25824: 29022: contig of 3198 bp in length
29023: 29123: gap of unknown length
29123: 31270: contig of 2148 bp in length
31271: 31370: gap of unknown length
31371: 35314: contig of 3944 bp in length
35315: 35414: gap of unknown length
35415: 37893: contig of 2479 bp in length
37894: 37993: gap of unknown length
37994: 41088: contig of 3095 bp in length
41089: 41188: gap of unknown length
41189: 43791: contig of 2603 bp in length
43792: 43891: gap of unknown length
43892: 46961: contig of 3070 bp in length
46962: 47061: gap of unknown length
47062: 50215: contig of 3154 bp in length
50216: 50315: gap of unknown length
50316: 52419: contig of 2104 bp in length
52420: 52519: gap of unknown length
52520: 55806: contig of 3287 bp in length
55807: 55906: gap of unknown length
55907: 59295: contig of 3389 bp in length
59296: 59395: gap of unknown length
59396: 61798: contig of 2403 bp in length
61799: 61898: gap of unknown length
61899: 64243: contig of 2345 bp in length
64244: 64343: gap of unknown length
64344: 67007: contig of 2664 bp in length
67008: 67107: gap of unknown length
67108: 68873: contig of 1766 bp in length
68874: 68974: 71449: contig of 2476 bp in length
71450: 71550: gap of unknown length
71550: 74329: contig of 2780 bp in length
74330: 74429: gap of unknown length
74430: 76070: contig of 1641 bp in length
76071: 76170: gap of unknown length
76171: 78425: contig of 2255 bp in length
78426: 78525: gap of unknown length
78526: 80194: contig of 1669 bp in length
80195: 80294: gap of unknown length
```

```
* 80295: 81542: contig of 1248 bp in length
* 81543: 81642: gap of unknown length
* 81643: 83636: contig of 1994 bp in length
* 83637: 83736: gap of unknown length
* 83737: 86415: contig of 2679 bp in length
* 86416: 86515: gap of unknown length
* 86516: 88344: contig of 1829 bp in length
* 88345: 88444: gap of unknown length
* 88445: 90189: contig of 1745 bp in length
* 90190: 90289: gap of unknown length
* 90290: 92121: contig of 1832 bp in length
* 92122: 92221: gap of unknown length
* 92222: 93934: contig of 1713 bp in length
* 93935: 94034: gap of unknown length
* 94035: 96058: contig of 2024 bp in length
* 96059: 96158: gap of unknown length
* 96159: 97881: contig of 1723 bp in length
* 97882: 97981: gap of unknown length
* 97982: 99055: contig of 1074 bp in length
* 99056: 99155: gap of unknown length
* 99156: 100698: contig of 1543 bp in length
* 100699: 100798: gap of unknown length
* 100799: 102147: contig of 1349 bp in length
* 102148: 102247: gap of unknown length
* 102248: 105281: contig of 3044 bp in length
* 105292: 105391: gap of unknown length
* 105392: 107562: contig of 2171 bp in length
* 107563: 107662: gap of unknown length
* 107663: 108956: contig of 1294 bp in length
* 108957: 109056: gap of unknown length
* 109057: 110501: contig of 1445 bp in length
* 110502: 110601: gap of unknown length
* 110602: 112705: contig of 2104 bp in length
* 112706: 112805: gap of unknown length
* 112806: 114186: contig of 1381 bp in length
* 114187: 114286: gap of unknown length
* 114287: 116172: contig of 1886 bp in length
* 116173: 116272: gap of unknown length
* 116273: 117323: contig of 1051 bp in length
* 117324: 117423: gap of unknown length
* 117424: 119313: contig of 1890 bp in length
* 119314: 119413: gap of unknown length
* 119414: 120518: contig of 1105 bp in length
* 120519: 120619: gap of unknown length
* 120619: 122209: contig of 1591 bp in length
* 122210: 122309: gap of unknown length
* 122310: 123422: contig of 1113 bp in length
* 123423: 123522: gap of unknown length
* 123523: 125754: contig of 2232 bp in length
* 125755: 125854: gap of unknown length
* 125855: 127342: contig of 1488 bp in length
* 127343: 127442: gap of unknown length
* 127443: 129015: contig of 1573 bp in length
* 129016: 129115: gap of unknown length
* 129116: 130755: contig of 1640 bp in length
* 130756: 130855: gap of unknown length
* 130856: 132296: contig of 1441 bp in length
* 132297: 132397: gap of unknown length
* 132397: 133904: contig of 1508 bp in length
* 133905: 134004: gap of unknown length
* 134005: 135616: contig of 1612 bp in length
* 135617: 135716: gap of unknown length
```

```
alignment_scores:
    Quality: 1777.00      Length: 408
    Ratio: 4.664          Gaps: 5
    Percent Similarity: 93.382    Percent Identity: 89.216
```

alignment_block:

US-09-811-045A-1 x AC095130 ..

Align seg 1/1 to: AC095130 from: 1 to: 153479

659 SerProPheThrArgGlyArgGlyAspGlyValAlaGlyGlnSerH 675
 107681 AGTACTTATCAAGAGGTTCCAGAGAGACTATGTCGACCAAGTCA 107730
 675 SARAGSARAGSLeuGlyGlyAsnTyrProGlnLysLeuSerThrArg 692
 107731 TGGAAATAGAAATCTAGTGGCACTATCCAGAAAAAGATTGACAGAG 107780
 692 SPSerHisAsnAlaLysAspAsnProLysSerLysGlnLysGlnSerGlu 708
 107781 ATAGTCACAAATCCAAAGATATCCAAAGTCCAAAGAGAAAGAGAGTGA 107830
 709 AsnValProGlnLysAspGlyLysGlnLysAsnLysHisLysArgLys 725
 107831 AATGTCCAGGAGCGCAAGAGAAACAGCATAGAAACACAGGAAAGC 107880
 725 GATGASnGlnLysGlyGlyGlnLysSerLysPheLeuAsnProGlnL 742
 107881 AAGA.....AAGGGGAGAGAGGTGAGAGCTTCGAAACCCAGAC 107921
 742 euLeuGlnThrSerArgLysCysArgGlySerSerGlyTleAspGlnThr 758
 107922 TACTGGAAATCTTAGAAATCCAGGGAATCCTCAGGAGATCGATGAACA 107971
 759 LysThrAspThrLeuPheValLeuProSerArgAspAspAlaThrPro 775
 107972 AAGACGAGATGACTGTTGTCTCCAGACAGAGATGCTACACTCT 108021
 775 IArgAspGlnPrometAspAlaGlnSerLleThrPheLysSerValSer 792
 108022 TGGGATGAAACCAATGATGAGAGATCATCTCAATCAATGATGTCG 108071
 792 SPlysAspLysArgGlnLysAspLysProLysValLysSerAspLysThr 808
 108072 AAAAAACAAAGAGAGAAAGATAAACCAAAAGTAAAGTACAAAGACC 108121
 809 LysArgLysSerAspLysSerAlaThrAlaLysAspAsnValLeuLys 825
 108122 AAACGGAAAAAGATGGTCTCTACAGCCCAAGAAAAAGTTTAA 108171
 825 SPSerLysGlyProGlnLysValAspLysAspArgLysSerP 842
 108172 ACCTCTAAAGAGACCCCAAGAGATAGTACGACCTGAAAAATCTC 108221
 842 rArgSerGlnProProLeuLysLysAlaLysGlnLysAlaThrLysIle 858
 108222 CTGGGCTGTAACCTCTCTCAAAAAAGAGAGAGGCTCAAAAGGTC 108271
 859 ASPSerValLysProSerSerSerSerGlnLysAspGlnLysValThrG 875
 108272 GACTCTGTAACCT...TCTCATCTCAGAAAGATGAGAGTCACTG 108318
 875 YThrProArgLysAlaHisSerLysSerAlaLysAsp.ThrArgArgGln 891
 108319 AACCCCTAGAAAAAGCCCACTCCAAATCTCAAAAGAACACAGAGGCAA 108368
 892 SerGlnProArgThrArgArgSerLysArgThrValProLysThrSerSe 908
 108369 ACCCAGTCAAAAGAGAGAGCAAAAAAGAGACTGTTCCAAAGACATCAAG 108418
 908 rGlnLysSerGlnProValArgThrArgArgProArgSerLysArgLysI 925
 108419 TCAGAAAAACCCAGCCAGTAGAGAGAGAAAGCCCAAGAACCTGAGAAAA 108468
 925 IAsnTyrLeuIleAlaArgGlnLysAsnGlnLysGlnLysArgLys..L 941
 108469 TAAACTACTTGAAGAGAGAGAGAAAGAAAGAAAGAAAGAAAGAGAGAG 108518
 941 YSerSerLysLysAspPheGlnLysSerSerSerMetLysLysSerLysVal 957
 108519 AAGAGTGA.....GATTTTGAAGTCTCTTCAATGAAAAATCTTAAAGTG 108562
 958 GlnGlyThrGlnLysValLysProSerProLysArgLysMetGlnLys 974

108563 GAAGGACGAGAAATAGTGAACCATCACCAAGCCGAAATGCAAGTGA 108612
 974 pValGlnLysLeuGlnArgThrProGlnLysAspLysIleAlaSerSerT 991
 108613 TGTGAAAGAGCTGGAAGAGACCCCAAGAAAGCAAGATGCATATCA 108662
 991 hrThrProAlaLysLysLysLysLeuAsnArgGlnThrGlnLysLysIle 1007
 108663 TTACCCACCCCAAAAAATTAACATCAATGAGAAACTGAAAAAAATT 108712
 1008 GlysAsnAlaGlnAsnAlaSerThrThrLysGlnProSerGlnLysLeuI 1024
 108713 GGAATGCGAAAAATCCATTACTACAAAAGAACCTATGAAAAACTGGA 108762
 1024 uSerThrSerSerLysLysLysGlnLysValLysGlyLysAlaLys 1041
 108763 GTCACATCTAACAATAATCAACAGAGAAAAAGTCAGAAAGGCCAGAC 108812
 1041 rGlnLysValAlaGlySerGlnLysSerSerSerThrLeuValAspLysThr 1057
 108813 GGAAATGACTGGAGACAGAGAGCCCAAGCTCCAGCTTGATGACTACAC 108862
 1058 SerThrSerThrGlyGly 1064
 108863 CAGGATGAGAGCCCGGTGG 108883
 seq_name: gb_pr:AK026954
 seq_documentation_block:
 LOCUS AK026954 2320 bp mRNA linear PRI 29-SEP-2000
 DEFINITION Homo sapiens cDNA: FLJ23301 fis, clone HEP11120.
 ACCESSION AK026954
 VERSION AK026954.1 GI:10439935
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone_11b:HEP clone:HEP11120.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (sites)
 Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Ohtani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S. NEDO human cDNA sequencing project
 TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished (2000)
 REFERENCE 2 (bases 1 to 2320)
 AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
 TITLE Direct Submission
 JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@elms.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction; 5'- & 3'-end one pass sequencing; Departent of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
 FEATURES
 source
 1..2320
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HEP11120"
 /cell_line="HEPG2"
 /cell_type="hepatoma"
 /clone_11b="HEP"
 /note="Cloning vector pME18SF13"
 3..1889
 /note="unnamed protein product"

[illegible]

BASE COUNT	957 a	397 c	462 g	504 t
ORIGIN				

alignment_scores:

Quality:	1750.00	Length:	455
Ratio:	4.187	Gaps:	3
Percent Similarity:	91.868	Percent Identity:	76.044

```
alignment_block:
```

US-09-811-045A-1 x AK026954 .

Align seg 1/1 to: AK026954 from: 1 to: 2320

```

952 MetLysIleSerLeuValGlnGlyThrGluLeuValLysProSerProly 968
    |||||
3  ATGAAATTCCTGAAACTCTGAAAGTGCAGTAATATGGAACCATCCACAA 52
968 sArGlyMetGlnGlyAspValGlnLysLeuGlnArgThrProGluLysA 985
    |||||
53 GCGCAAAATGGAACCTGATACAGAAAAATGGATRGACCCCTGAAAGG 102
985 sPylsIleAlaSerSerThrThrProAlaLysLysIleLysLeuAsnArg 1001
    |||||
103 ACAAAAT...TCTTAAGTGGCCAGCCAAAAAAATCAACTCAACGA 149
1002 GluThrGlyLysLysIleGlyAsnGlnGlnAlaSerThrThrLysGln 1018
150 GAAACTGGGACAAATTTGGAGTGCACAGAAATATATCAACACAAAGA 199
1018 uProSerGlnLysLeuGlnSerThrSerSerLysIleLysGlnLysV 1035
200 ACCCTCTGAAAAATTTGGAGTCAACTCTAGCAAACTTAACAGAAAG 249
1035 AlLysGlyLysAlaLysArgLysValAlaGlySerGlnGlySerSer 1051
    |||||
250 TCAAAAGAAAGGTCTACAGCAAAAGTACTGGCAACGAAGATCCAGCTCA 299
1052 ThrLeuValAspTyrThrSerThrSerSerThrGlyGlySerProValAr 1068
300 ACTCTGGTGGAATTACACCAAGTACGAGCTCACTGGAGGCAAGTCTGTG 349
1068 GlySerGlnGlnLysThrAspThrLysArgThrValIleLysThrMetG 1085
350 GAAATCTGAGAAAAAAGAGATACCAACGCAACTGTGATTAACAGATGG 399
1085 LysGlyThrAsnAsnAspAsnThrAlaProAlaGlnAspValIleLeuMet 1101
400 AGCAATATATATATACATATACCGGCCACGCTGAAGATGTTATCTTATG 449
1102 IleGlnValProGlnSerLysThrAspLysAspAspPheGlnSerGlnI 1118
450 ATTCAAGTTCTCCATCAATCCAAATGGATTAAGATGACTTTGAATCGAAG 499
1118 uGlnAspValLysThrThrGlnProIleGlnSerValGlyLysProSer 1135
500 ACAAATGTTAAATCCACACACCTATCTACAGTGTAGAAAAACCTGCTGA 549

```

```

1135 erllellelysaenvalThrThrlysproserAlaThrAlaLystrThr 1151
1136 |::::::::::::::::::::::::::::::::::::|
550 GTGTTTAAAAAATGTTATGTTACAAAGCCATCAAAATATATGTAAGTATCT 599
1152 GluLysGluSerGluGlnProGluLysGluGlnLysLeuProLysGluAl 1168
1153 |::::::::::::::::|
600 GAGAAAGAAAGTAGCCATCCGAGAAAATTCCAGAAATTCACCAAGGAGCT 649
1168 aserHisGluLeuLeuGlnHisGluLeuAlaArgSerLysGluSerAlas 1185
1169 |::::::::::::::::|
650 GAGCCATGAAGATCATACACATGAGGTAAAGTTCAAAAAAACCTCGAT 699
1185 eSerGluLysGluArgAlaLysAspArgGluHisSerGlySerGlyLys 1201
1186 |::::::::::::::::|
700 CTAGTCAAAAAGGAGAAACCAAGATGACACTTATTTCAGTCTTGAAAG 749
1202 AspAsnProAspLysArgLysSerGlyAlaGlnProAspLysGluSerTh 1218
1203 |::::::::::::::::|
750 GAGAACCTGAAAGAGAGAAACAGACACTCAGCCAGAAAGAGAGTAA 799
1218 rValAspArgLeuSerGluGlnGlyHisPheLysThrLeuSerGlnSer 1235
1219 |::::::::::::::::|
800 TTGGACCGCTGTAATGAAACAGAAATTTTAAAAAGCTGCTCAACTT 849
1235 eArgLysGluThrArgThrSerGluLysHisGluSerValArgGlySerSer 1251
1236 |::::::::::::::::|
850 CCAAAGAGCGTAGACACTCAGATTAACATGATTTCCACTCGCTCTCTCA 899
1252 AsnLysAspPheThrProGlyArgAspLysLysValAspTryAspSerAr 1268
1253 |::::::::::::::::|
900 AATTAACACTTCACTCCCAATAGAGCAAAAAAATCAGTACTGTGACACAG 949
1268 gaSPTrSerSerSerLysArgArgAspGluArgGluLeuLeuAlaArg 1285
1269 |::::::::::::::::|
950 AGAGTATTCAAGTTCCAACGCTGAGAGTGAAGAAATAAGATTAACAAGAC 999
1285 rGlyAspSerProProArgGlyLysGluSerLeuSerGlyGlnLysSer 1301
1286 |::::::::::::::::|
1000 GAAAGACTCTCTCTTCGGAATTAAGTTCTGCATCTGAGACAGAAAT 1049
1302 LysLeuArgGluLuarGAspLeuProLysLysGly..AlaGluSerLys 1317
1303 |::::::::::::::::|
1050 AAACCAAGGAGAAAGAGAGATTTCCCTAAAAAAGAACAGGAGATTTCCA 1099
1317 sLysSerAsnSerSerProProArgAspLysLysProHisAspHisLysA 1334
1318 |::::::::::::::::|
1100 AAAAAGATTAATTCTGCTCCCTCAGAGACAGAAAACTCATGATACAAAG 1149
1334 laProTyrGluThrLysArgProCysGluGluThrLysProValAspLys 1350
1335 |::::::::::::::::|
1150 CCACTTATGTACTAAACGGCCAAATGACAGACAAAATCTGTGATGATAA 1199
1351 AsnSerGlyLysGluArgGlyLysHisAlaAlaGluLuarGAsnGlyLys 1367
1352 |::::::::::::::::|
1200 AATCTTGTGAAGACTGTGAGAACCAAGTATTAACACCAAGGAACAATAA 1249
1367 sGluSerSerGly..AlaAsnCysHisValTyrLeuThrArgGlnThrLeu 1383
1368 |::::::::::::::::|
1250 AGAGTCAAGGCGCAATAAAGTACTTTATATACCTTAACCCACCAAGACAC 1299
1384 ProTTrpArgArgSerTrpLeuLeuGlyArgTrpArgAlaAlaProSerSe 1400
1385 |::::::::::::|
1300 AGGTGTGAAAAAGACAATTAATCTGGCAAAATTGCAAGAGTACTGTCAAG 1349
1400 rArgAsnProSer 1404
1401 |::::::::::::|
1350 CCTAAACCCAGCT 1362
seq_name: qb_htg:AL670951

```

seq_name: gkb_htg:AL670951

seq_location	chr20p11.2
locus	AL670951
definition	Mus musculus chromosome 2 clone RP23-6N18, *** SEQUENCING IN
size	204937 bp
map	linear HTG 08-FEB-2002

ACCESSION	AL670951	PROGRESS ***, in unordered pieces.
VERSION	AL670951.4	GI:18643842
KEYWORDS	HTGS; HTGS_PASEL1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus	
AUTHORS	1 (sites)	
TITLE	Bates,K.	
JOURNAL	Direct Submission	
	Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,	
	Cambridgeshire, CB10 1SA, UK. E-mail enquiries:	
	humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk	
	On Feb 10, 2002 this sequence version replaced gi:18491458.	
COMMENT	----- Genome Center	
	Center: Wellcome Trust Sanger Institute	
	Center code: SC	
	Web site: http://www.sanger.ac.uk	
	Contact: humquery@sanger.ac.uk	
	----- Project Information	
	Center project name: bM6N18	
	----- Summary Statistics	
	Assembly program: XGAP4; version 4.5	
	Sequencing vector: plasmid; L08752; 100% of reads	
	Chemistry: Dye-terminator Big Dye; 100% of reads	
	Consensus quality: 203030 bases at least Q40	
	Consensus quality: 203337 bases at least Q30	
	Consensus quality: 203532 bases at least Q20	
	Insert size: 203737; sum-of-contigs	
	Insert size: 208368; 7.5% error; agarose-fp	
	Quality coverage: 13.20x in Q20 bases; sum-of-configs Quality	
	coverage: 13.09x in Q20 bases; agarose-fp	

	* NOTE: This is a 'working draft' sequence.	
	* This record will be updated with the finished sequence	
	* as soon as it is available and the accession number will	
	* be preserved.	
FEATURES		
source	1. 204937	Location/Qualifiers
	/organism="Mus musculus"	
	/db_xref="taxon:10090"	
	/chromosome="2"	
	/clone="RP23-6N18"	
	/clone_1lb="RPc1-23"	
	1. 26114	
misc_feature	/note="assembly_fragment:04643	
	clone_end:77	
	vector_side:left"	
misc_feature	26215..35512	
	/note="assembly_fragment:05542	
	fragment_chain:1"	
misc_feature	35613..64166	
	/note="assembly_fragment:04831	
	fragment_chain:1"	
misc_feature	64267..69715	
	/note="assembly_fragment:05492	
	fragment_chain:1"	
misc_feature	69816..101415	
	/note="assembly_fragment:01175	
	fragment_chain:2"	
misc_feature	101516..109750	
	/note="assembly_fragment:02972	
	fragment_chain:2"	
misc_feature	109851..121829	
	/note="assembly_fragment:01319	
	fragment_chain:3"	
misc_feature	121930..131947	
	/note="assembly_fragment:06136	
	fragment_chain:3"	
misc_feature	132048..139628	
	/note="assembly_fragment:03401	
	fragment_chain:3"	

misc_feature	139729..149923	/note="assembly-fragment:02273
misc_feature	150024..165181	/fragment_chain:3*
misc_feature	165282..202703	/note="assembly-fragment:06020
misc_feature	165282..202703	/fragment_chain:3*
misc_feature	165282..202703	/note="assembly-fragment:03330
misc_feature	165282..202703	/fragment_chain:3*
misc_feature	165282..202703	/note="assembly-fragment:05292
misc_feature	165282..202703	/fragment_chain:3
misc_feature	165282..202703	clone_end:SP6
misc_feature	165282..202703	vector_slide:"right"
BASE COUNT	52741 a 47028 c 47279 g 56686 t 1203 others	
ORIGIN		
alignment_scores:	Quality: 1585.50	Length: 503
Percent Similarity:	Ratio: 4.129	Gaps: 10
Percent Similarity:	76.342	Percent Identity: 66.203
alignment_block:	US-09-811-045A-1 x AL670951/rev ..	
Align seg 1/1 to reverse of: AL670951 from: 1 to: 204937		
237	SettleServalHisSerLulysSerAspGlyProPheArgAspSerAs	253
23664	AGTTTATATATCCATTCACGAAATATCAGTATGACCTTTCGGATTCGA	23615
253	PAenLYLeuLeuProAlaAlaAlaLeuThSerGluHisSerLYsGLYA	270
23614	TATCAATATATATACGAGCTGCCACCTTATCATCAGACATTTAAAGGAG	23565
270	laserSerLleAlaIleThraAlaLeuMetGluLysGLyValProgly	286
23564	CCCTTCATTTGGCTGTACTGCTCTTATGAGAAAGTG..TGCCAGGC	23516
287	ThSerProTrAspSerLlePheVal..GlyGlnSerLeuLeuHISGLY	303
23515	GCCAGTCCTTGAAACGCCACCTTTGTTGGAGACATTTATACACGAC	23466
303	lLeuLleProThrThrGlyProValArgLLeaSnAlaAlaArgProgly	319
23465	AGTTTATCCCAACATGCGTCCAGTAAATGATGCTGCTTCCAGCT	23416
320	GlyGLYArgProGlyTrpGluHisSerAenLYsLeuGLYTYrLleuVal	336
23415	GGTGGTCAACAGGTTGGGAGACATTCACCAAGCATGGGCTTGTGTTG	23366
336	rProProGlnGlnLLeArgArgLYGluArgSerCYsTYrArgSerLle	353
23365	CCACACACAGAAATATAGAGAGAACAAAGAGCGTTTACAGAGATTA	23316
353	SnArgLYArgHISHisSerGluArgSerGlnArgThrGlnSerProser	369
23315	AAGCGTGGCAACACTACAGACATCATCAGAGAGGATCAAGGCCGTC	23266
370	LeuProAlaThProCysPheValProValProProProLeuLysPr	386
23265	CGACCGACAACTCCATCTTTTGGCTGTGCCACCGCTCATGCAAC	23216
386	oProProProHISThLeuProLeuProProLYValProProProGlnP	403
23215	T.....CTTCCCTTCTCTAGTATATACCTCCACAGT	23181
403	heserProGlnPheProSerSerGlnProProThAlaAsLYrSerVal	419
23180	TTTCTCTCTAGTTTCTCTGCTGCTACCGACGACGAGATATAGTGG	23131
420	ProProProGlyPheProProAlaProAlaSnLleSerThAlCYsPh	436

```

|||||
23130 CTTCTCCAGAGATCCACGAGCTCTGCCAATATATCAACAGCTGGGT 23081
436 eSerProgluylalProthrAlaHisSerAnthrMetProthrThrGlna 453
|||||
23080 TTCTCAGTAGTGCAGACTGCCCTTCAAAATACATCTATTAACACAG 23031
453 laProleuLeuSerArgGluGluPheThrArgGluGlnAsnAspLysGly 469
|||||
23030 CACTTCTTGTTCACAGGAAGATTTCTTAGAGGCAACAACACTCTAAG 22981
470 ArgGluSerLysPheProtyrSerGlySerSerTySerArgSerSerTy 486
::: |||||
22980 GAACGCTCTAATTTCCCTACAGTGGGTCAATCGTATTCAGAAATTCATA 22931
486 rThAspSer.SerGlnGlyLeuAlaGlnHisThrHisAlaLeuThrLeu 502
|||||
22930 CACTGACTCAAAAGTCAAGTGTGGCTCAACACAGCTACACTTACTCTA 22881
503 SerProSerAlaAlaHisThrLeuAspLeuHisAspHisProHisSpr 519
|||||
22880 AGTCTCTCAGAGCTCACACTCTGCTCTTCTTACAGATCATCTCCCATCC 22831
519 oProGluGluAlaGluAlaArgSerAlaMetIleValHisMetProAspL 536
|||||
22830 TCCAGAGAGAGCAGAGCAAGATCTGCATGACATCATCATGTCAGATC 22781
536 euMetAspLleAla..... 540
|||||
22780 TCATGATATTCACAGATCTAGTCAAGTCCCTCCCTGATGTCATGGA 22731
541 .....HisAlaArgSerArgSerProProtyrArgArgTyArgSerAr 555
|||||
22730 TATCAACAGATCTAGTCAAGTCCCTCCCTATATGACATATCACTCATG 22681
555 gSerArgSerProProGluPheArgGlyGlnSerProThrLysArgAsnV 572
|||||
22680 GTCCAGATCTCCCTCAGCATTTAAGAACACTCTCCCACTAACTCATATG 22631
572 alProArgGluGluLysGluArgGluTyrPheAsnArgTyArgLysVal 588
|||||
22630 TCCCTCAAGGA..... 22620
589 ProProProtyrAspLleLysAlaTyrTyrGlyArgSerValAspPheAr 605
22620 ..... 22620
605 gAspProPheGluLysGluArgTyArgGluTyArgLysArgLysTyArgG 622
|||||
22619 .....GAGAGAGAGACGCTACCAAGAAATGGAGAGCAACACAGG 22580
622 LuTrpTyArgLysTyArgTyArgLysAlaValGlyAlaGlnProAr 638
|||||
22579 GGTGTATGACAAAT..AGGCAAACGGCAGACAGTACAGACAACTTACG 22533
638 gProSerAlaAsnArgLysAspPheSerProGluArgLeuProleuA 655
|||||
22532 ACCCTAGGCCAACAGAGAGTCTTTCTCCAGAGAGAACTTACCTCTTA 22483
655 snIleArgAsnSerProPheThrArgGlyArgArgGluAspTyrAlaAla 671
|||||
22482 ATATCATTAACCCACCTTCCCAAGAGGGCACAGAGAAAGACTATGTGCT 22433
672 GlnGlnSerHisArgAsnArgAsnLeuGlyLysAsnTyProGluLysLe 688
|||||
22432 GGGCAAAATCTGTAATCA.....GGAGCAACTACCCAGAAATATT 22389
688 uSerThrArgAspSerHisAsnAlaLysAspAsnPro.Lys.....Ser 702
|||||
22388 TTCAACAAGACAGACTCACAATGCAAAAGATPACCAAAAGATATACACA 22339
703 LysGluLysGluSerGluAsnValProGluAspGlyLysGlyAsnLysHis 719
|||||

```

```

22338 AAAGAGAAAGAAAGTGAATGTTCCAGAGACAGCAAGAAAGAAACAGCA 22289
719 s 719
22288 T 22288
seq_name: gb_pr:HSN802662
seq_documentation_block:
LOCUS HSN802662 2026 bp mRNA linear PRI 18-JUN-2000
DEFINITION Homo sapiens mRNA: cDNA DKFZp761B2423 (from clone DKFZp761B2423).
ACCESSION AL359564
VERSION AL359564.1 GI:8655620
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2026)
REFERENCE
Bloecker,H., Boecker,M., Brandt,P., Mewes,H.W., Weil,B. and
Wiemann,S.
Direct Submission
Submitted (15-JUN-2000) MIPS, Am Klopfersplitz 18a, D-82152
Martinsried, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp761B2423) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
Location/Qualifiers
1. 2026
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp761B2423"
/issue_type="amygdala"
/clone_lib="761 (synonym: hamy2). Vector pSport1; host
DH10B; sites NCI + SalI"
/dev_stage="adult"
1..1592
/gene="DKFZp761B2423"
<1..1592
/gene="DKFZp761B2423"
/note="similarity to p53-associated cellular protein PACT
(Mus musculus)"
/codon_start=3
/product="hypothetical protein"
/protein_id="CA94869.1"
/db_xref="GI:8655621"
/translation="TLDVSTSTSGSPVPRKSEKTDKRTVYIKTMEYNNNDNTAPA
EDVITIMIOVPSKWDKDFSEEDYKSTQPISSVQKPAVYIKNVTKSNLYKYEK
ESPEPRKIKTKFDVSHETIOHNVKSSKNSASEKGTDRDYSLEKNEPKRKST
QEPKESNDRNLNEQGNFKLSQSKARPSDKHDSRASNNKDPNRPKRDYDRE
YSSSKRDRKNEELTRRKDPSRKNDASQKPKPREKLPKRGTDSSKSSNSPDR
RKPHDKATYDTRPNEETKSYDKNCKDKREKHLVLEARNKSSGKLLYLNPETO
VERKEITGQIDKSTYKPKPOLSHSRSLSDLTRENDDEAFEDYNSDESNSVKEE
ESSGNSKDKKIVKAKESLDTAIVGVGISRNQSHSPSPSPSPSPSPSGSQTRS
HSSASASASQSKKKKKKKKKKKKKKKKKKKHAGTEVELEKSKKKKKKKSK
KKNDKEKEKEDQKYSYTV"
1983..1988
polya_signal
polya_site
BASE COUNT 820 a 344 c 405 g 457 t
ORIGIN
alignment_scores:
Quality: 1353.50 Length: 355
Ratio: 4.190 Gaps: 2
Percent Similarity: 90.986 Percent Identity: 74.648

```



```

seq.document.ac095130.      153479 bp   DNA   linear   HTG-20-DEC-2001
LOCUS      AC095130
DEFINITION Rattus norvegicus clone CH230-8F21, *** SEQUENCING IN PROGRESS ***,
            68 unordered pieces.
ACCESSION  AC095130
VERSION    AC095130.2 GI:17942059
KEYWORDS   HTG: HTGS_PHASE1.
SOURCE     Norway rat.
            Rattus norvegicus
            Eutheria; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;

```

REFERENCE	AUTHORS
1	(bases 1 to 153479)
REFERENCE	Katus.
1	(bases 1 to 153479)
REFERENCE	Munzy,D.M., Adams,C., Adio-Oduola,B., Ali-oshan,F.R., Allen,C.,
REFERENCE	Albrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbarta,J.,
REFERENCE	Benton,J., Blmage,K., Blankenburg,K., Bonndt,D., Bouch,J.,
REFERENCE	Boyle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Bulhay,C.,
REFERENCE	Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
REFERENCE	Cartier,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
REFERENCE	Chen,Z., Chowhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
REFERENCE	Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
REFERENCE	Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
REFERENCE	Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
REFERENCE	Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
REFERENCE	Elhaj,C., Escotto,M., Falis,T., Ferraguto,D., Flagg,N., Ford,J.,
REFERENCE	Foster,P., Frintz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
REFERENCE	Garrar,N., Gill,R., Gorrell,J.H., Guenara,W., Guarnine,P., Hale,S.,
REFERENCE	Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
REFERENCE	Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
REFERENCE	Hollins,B., Homs,I.F., Howard,S., Huber,J., Huly,S., Hume,J.,
REFERENCE	Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
REFERENCE	Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korval,J.,
REFERENCE	Kover,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
REFERENCE	Lewis,L., Li,T., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
REFERENCE	Louised,H., Lozado,R.D., Lu,X., Lucier,A., Lucier,R., Luna,R.,
REFERENCE	Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
REFERENCE	Marine,E., Massey,E., Mawliny,E., McLeod,M.P., Meador,M.,
REFERENCE	Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,
REFERENCE	Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
REFERENCE	Nguyen,A., Nguyen,S., Nguyen,N., Nickerson,E., Nwokewo,S.,
REFERENCE	Ogun,M., Okunodu,G., Oragunye,N., Oviado,R., Pace,A., Payton,B.,
REFERENCE	Peery,J., Ren,Y., Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M.,
REFERENCE	Quiles,M., Kren,Y., Scherer,S., Scott,G., Shen,H., Shoshartai,N.,
REFERENCE	Ruiz,S., Severy,G., Scherer,S., Scott,G., Shen,H., Shoshartai,N.,
REFERENCE	Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,N.,
REFERENCE	Stone,H., Sutton,A., Vatek,A., Tabori,P., Tamerisa,A., Thomas,N.,
REFERENCE	Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Theras,R.,
REFERENCE	Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R.,
REFERENCE	Walli,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
REFERENCE	Wallington,S., Williams,G., Williamson,S., Wlecyxg,R., Wooden,S.,
REFERENCE	Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
REFERENCE	Weinstock,G. and Gibbs,R.
REFERENCE	Direct Submssion
REFERENCE	2 (bases 1 to 153479)
REFERENCE	Morley,K.C.
REFERENCE	Direct Submssion
REFERENCE	Submitted (16-SEP-2001) Human Genome Sequencing Center, Departmenth
REFERENCE	of Molecular and Human Genetics, Baylor College of Medicine, One
REFERENCE	Baylor Plaza, Houston, TX 77030, USA
REFERENCE	On Dec 20, 2001 this sequence version replaced gf:15625684.
REFERENCE	----- Genome Center
REFERENCE	Center: Baylor College of Medicine
REFERENCE	Center code: BCM
REFERENCE	Web site: http://www.hgsc.bcm.tmc.edu/
REFERENCE	Contact: hgsc-help@bcm.tmc.edu
REFERENCE	----- Project Information
REFERENCE	Center project name: GC0M

```
Center clone name: CH230-8F21
----- Summary Statistics -----
Assembly Program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 121430 bases at least Q40
Consensus quality: 129669 bases at least Q30
Consensus quality: 136708 bases at least Q20
Estimated insert size: 119515; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 1.6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 68 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 6858: contig of 6858 bp in length
6859 6958: gap of unknown length
6959 12717: contig of 5759 bp in length
12718 12817: gap of unknown length
12818 16979: contig of 4162 bp in length
16980 17079: gap of unknown length
17080 21262: contig of 4183 bp in length
21263 21362: gap of unknown length
21363 25724: contig of 4362 bp in length
25725 25824: gap of unknown length
25825 29022: contig of 3198 bp in length
29023 29122: gap of unknown length
29123 31270: contig of 2148 bp in length
31271 31370: gap of unknown length
31371 35314: contig of 3944 bp in length
35315 37893: contig of 2479 bp in length
37894 37993: gap of unknown length
41088: contig of 3095 bp in length
41089 41188: gap of unknown length
41189 43791: contig of 2603 bp in length
43792 46961: contig of 3070 bp in length
46962 47061: gap of unknown length
50215: contig of 3154 bp in length
50216 50315: gap of unknown length
50316 52419: contig of 2104 bp in length
52420 52519: gap of unknown length
52520 55806: contig of 3287 bp in length
55807 55906: gap of unknown length
55907 59295: contig of 3389 bp in length
59296 59395: gap of unknown length
59396 61798: contig of 2403 bp in length
61799 61898: gap of unknown length
61899 64243: contig of 2345 bp in length
64244 64343: gap of unknown length
64344 67007: contig of 2664 bp in length
67008 67107: gap of unknown length
67108 68873: contig of 1766 bp in length
68874 68974: gap of unknown length
68974 71449: contig of 2476 bp in length
71450 71549: gap of unknown length
71550 74329: contig of 2780 bp in length
74330 74429: gap of unknown length
74430 76070: contig of 1641 bp in length
76071 76170: gap of unknown length
76171 78425: contig of 2255 bp in length
78426 78525: gap of unknown length
78526 80194: contig of 1669 bp in length
80195 80294: gap of unknown length
80295 81542: contig of 1248 bp in length
81543 81642: gap of unknown length
```

```
* 81643 83636: contig of 1994 bp in length
* 83637 83736: gap of unknown length
* 83737 86415: contig of 2679 bp in length
* 86416 86515: gap of unknown length
* 86516 88344: contig of 1829 bp in length
* 88345 88444: gap of unknown length
* 88445 90189: contig of 1745 bp in length
* 90190 90289: gap of unknown length
* 90290 92121: contig of 1832 bp in length
* 92122 92222: gap of unknown length
* 92223 93934: contig of 1713 bp in length
* 93935 94034: gap of unknown length
* 94035 96058: contig of 2024 bp in length
* 96059 96158: gap of unknown length
* 96159 97881: contig of 1723 bp in length
* 97882 97981: gap of unknown length
* 97982 99055: contig of 1074 bp in length
* 99056 99156: gap of unknown length
* 99157 100698: contig of 1843 bp in length
* 100699 100798: gap of unknown length
* 100799 102147: contig of 1349 bp in length
* 102148 102247: gap of unknown length
* 102248 105291: contig of 3044 bp in length
* 105292 105391: gap of unknown length
* 105392 107562: contig of 2171 bp in length
* 107563 107662: gap of unknown length
* 107663 108956: contig of 1294 bp in length
* 108957 109056: gap of unknown length
* 109057 110501: contig of 1445 bp in length
* 110502 112705: contig of 2104 bp in length
* 112706 112805: gap of unknown length
* 112806 114186: contig of 1381 bp in length
* 114187 114286: gap of unknown length
* 114287 116172: contig of 1886 bp in length
* 116173 116272: gap of unknown length
* 116273 117323: contig of 1051 bp in length
* 117324 117423: gap of unknown length
* 117424 119313: contig of 1890 bp in length
* 119314 119413: gap of unknown length
* 119414 120518: contig of 1105 bp in length
* 120519 120618: gap of unknown length
* 120619 122209: contig of 1591 bp in length
* 122210 122309: gap of unknown length
* 122310 123422: contig of 1113 bp in length
* 123423 123522: gap of unknown length
* 123523 125754: contig of 2232 bp in length
* 125755 125854: gap of unknown length
* 125855 127342: contig of 1488 bp in length
* 127343 127442: gap of unknown length
* 127443 129015: contig of 1573 bp in length
* 129016 129115: gap of unknown length
* 129116 130755: contig of 1640 bp in length
* 130756 130855: gap of unknown length
* 130856 132296: contig of 1441 bp in length
* 132297 132396: gap of unknown length
* 132397 133904: contig of 1508 bp in length
* 133905 134004: gap of unknown length
* 134005 135616: contig of 1612 bp in length
* 135617 135716: gap of unknown length
```

```
alignment_scores:      Quality: 907.00      Length: 201
                        Ratio: 4.903      Gaps: 1
Percent Similarity: 92.040      Percent Identity: 87.065

alignment_block:
US-09-811-045A-1 x AC095130/rev ..
```

Align seg 1/1 to reverse of: AC095130 from: 1 to: 153479

```
472 SerLysPheProTyrSerGlySerTyrSerArgSerSerTyrThrAs 488
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

```

109663 TCTAAATCTCCCTATAGTGGTCATGCTATTCAGAGATTCATACACTTA 109614
488 pser.SerGinglyLeuAlaGlnHisIleHisAlaLeuThrLeuSerPro 504
109613 CTAAAGTCAAGTGTGGCTCAACAGCTTCACGCTTACTCTCGGCTT 109564
505 SerAlaAlaHisThrLeuAspLeuHisAspHisProHisProHisProHis 521
109563 TCAGCGGCTCAGACTCTCGCTCTATTCACGGTCACCGCCCATACCCAGG 109514
521 uGuaIaGlnAlaArgSerAlaMetIleValHisMetProAspLeuMetA 538
109513 AGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 109464
538 spIleAlaHisAlaArgSerArgSerProProTyArgArgArgSer 554
109463 ATACCAT.CGATCTAGTCAAGGTCACCTCCCTATACAGATACCACTCA 109415
555 ArgSerArgSerProProGluPheArgGlyGlnSerProThrLysArgAs 571
109414 CGGTCGAGATCTCTCAAGCATTTAGGGGACAGCTCCCAACGCAA 109365
571 nValProArgGluGluLysGluArgGluTyArgPheAsnArgTyArgGlu 588
109364 TGTACTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 109315
588 alProProProTyArgSpIleLysAlaTyTyTyTyTyTyTyTyTyTyTyTy 604
109314 TTCACCCCTTATGATATCAACACCTATATAGCGGAGAGTGCTGACTTT 109265
605 ArgAspProPheGluLysGluArgTyArgGluTyArgLysArgLysArg 621
109264 AGAGACCCATTTGAGAAAGACGCTACCGGAGATGAGAGAGAAATACG 109215
621 gGluTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 638
109214 AGACTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCT 109165
538 rGProSerAlaAsnArgGluAspPheSerProGluArgLeuLeuProLeu 654
109164 GACCTTCAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 109115
655 AsnIleArgAsnSerProPheThrArgGlyArgArgGluAspTyArgAla 671
109114 AATATCAAGAAATTCACCTTACAAAGAGCGGAGAGAGAGAGAGAGAG 109065
671 a 671
109064 T 109064
seq_name: gb_hlg:AL450323
seq_documentation_block: 110015 bp DNA linear HTG 10-JUL-2001
LOCUS AL450323
DEFINITION Mus musculus chromosome 2 clone RP23-349016, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL450323
VERSION AL450323.4 GI:12539773
KEYWORDS HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 110015)
Pearce,A.
Direct Submission
Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jan 26, 2001 this sequence version replaced gi:11493349.
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC

```

```

Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouseghar.mrc.ac.uk
----- Project Information
Center project name: BM349016
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: pGAS4; version 10090
Chemistry: dye-terminator Big Dye; 99% of reads
Chemistry: dye-primer Big Dye; 0% of reads
Consensus quality: 109962 bases at least Q40
Consensus quality: 110005 bases at least Q30
Insert size: 119290; sum-of-contigs
Insert coverage: 16.70x in Q20 bases; sum-of-contigs quality
Quality coverage: 15.58x in Q20 bases; agarose-fp
coverage: 15.58x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..110015
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-349016"
/clone_1lb="RPCT-23"
1..110015
/note="assembly_fragment:01877
clone_end:SP6
vector_side:left
vector_end:right"
BASE COUNT 27835 a 27870 c 26415 g 27895 t
ORIGIN
alignment_scores:
Quality: 773.00 Length: 279
Ratio: 3.944 Gaps: 8
Percent Similarity: 70.251 Percent Identity: 60.573
alignment_block:
US-09-811-045A-1.x AL450323/rev ..
Align seg 1/1 to reverse of: AL450323 from: 1 to: 110015
460 GluPheTyArgGluGlnAsnAspLysGlyArgGluSerLysPheProTy 476
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110015 GAATCTTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 109966
476 rSerGlySerSerTySerArgSerSerTyTyThrAspSer.SerGingly 492
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109965 CAGTGGGTGATCGATATCCCAAAATTCATCAGTCAAGAGAGAGAGAG 109916
493 LeuAlaGlnHisIleHisAlaLeuThrLeuSerProSerAlaAlaHisTh 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109915 CTGGCTCAACAGACAGACACCTTACTTAAGTCTTCAGCAGCTCACAC 109866
509 rLeuAspLeuLeuHisAspHisProHisProGluGluAlaAla 526
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109865 TCTCGCTCTTCTTCAGATCAATCCCATCTCCAGAGAGAGAGAGAGCA 109816
526 rGSerAlaMetIleValHisMetProAspLeuMetAspIleAla..... 540
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109815 GATCTGCAATGACATTCATGTCAGATCTTCAGATATCAGATATCA 109766
541 .....HisAlaArgSerArg 545
109765 GATCAAGTCCCTCCCTTGAATTCATGATATCAACATCTAGGTCAAG 109716
545 gSerProProTyArgArgTyArgSerArgSerArgSerProProGluP 562

```

```
|||||.....|
109715 GTCCCTCCCTATAGACATACCACATGTCAGATCTCTCCAGCAT 109666
562 heArgGlyGlnSerProThrIysArgAsnValProArgGluGluLysGlu 578
|||||.....|
109665 TTAAGGAACAGTCTCCACTAAATCAATGCTCAAGA..... 109625
579 ArgGluTyrPheAsnArgTyrArgGluValProProProTyrAspIleIy 595
109625 ..... 109625
595 sAlaTyrTyrGlyArgSerValAspPheArgAspPropheGluLysGlu 612
109624 .....GAGAGAGAGC 109615
612 IgTyrArgGluTyrGluArgLysTyrArgGluTyrTyrGluLysTyrTyr 628
|||||.....|
109614 GCTACCAGGAATGGAGAGAGACACAGGGCGGTATGAGAT...AGG 109568
629 .LysGlyTyrAlaValGlyAlaGlnProArgProSerAlaAsnArgGlu 645
|||||.....|
109567 CAACGGCAGCAGTAGAAGCACTTCTAGACCTAGGCCAAGAGAGAG 109518
645 sPheSerProGluArgLeuLeuProLeuAsnIleArgAsnSerProphe 661
|||||.....|
109517 TCTTTCTCCAGAGAACTTACTCTAATATCATATAACCCACCTTC 109468
662 ThrArgGlyArgArgGluAspTyrAlaAlaGlyGlnSerHisArgAsnAr 678
|||||.....|
109467 CCAGAGGGCCAGAGAACTATGTCTGGCAAGTCGTAGAAATCA 109418
678 gAsnLeuGlyIysAsnTyrProGluLysLeuSerThrArgAspSerHis 695
109417 A.....GGAGCAACTACCCAGAAAGTTTCAACAGACACAGTCACA 109374
695 sAlaLysAspAsnPro.Lys.....SerLysGluLysGluSerGluAs 709
109373 ATGCAAAAGATATATCCAAAGAGATTAACCAAAAGAGAAAGTAGANA 109324
709 nValProGlyAspGlyLysGlyAsnLysHis 719
|||||.....|
109323 TGTTCCAGGAGACAGCAAGAAACACAT 109293
```

U.S. PAGE BLANK (USPTO)

OM of: US-09-811-045a-1 to: N.Geneseq_032802.* out_format : pfs
Date: Sep 12, 2002 5:17 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame-p2n.model -DEV=xlh
-O=/cgr2.1/USPTO.spool/US09811045/rnatc.12092002.132243.6018/app-query.fasta.1.1485
-DB=N.Geneseq_032802 -QPM=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsme2
-TRANS=human0.cdi -LIST=45 -DOCALIGN=200 -PHR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFM=pfis
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US09811045.@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-811-045a-1
Query length: 1404
Database: N.Geneseq_032802.*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 327.370000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA160201 +	5635.50	4071.63	1.9e-218	4	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA158415 +	5381.00	3888.76	2.9e-208	4	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH31144 +	4251.00	3073.38	7.7e-163	35	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH28668 +	4088.50	2944.16	1.2e-155	1	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAOC7601 +	1838.50	1334.77	5.3e-06	18	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH33151 +	1122.50	821.83	2.0e-37	3	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AB105843 +	544.00	393.43	1.4e-13	7	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AB105843 +	530.50	378.95	7.2e-13	7	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AB105842 +	528.00	398.09	9.9e-14	4	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA182535 +	477.00	340.95	1.2e-10	67	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV21511 +	451.00	339.19	1.5e-10	68	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH28296 +	413.50	302.24	1.7e-08	2	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AB105592 +	401.00	278.54	3.6e-07	1	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH1795 +	400.50	283.50	1.9e-07	9	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH81756 -	389.00	267.82	1.4e-06	2	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AA13213 -	374.50	267.34	1.5e-06	63	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH23695 +	374.00	264.28	2.3e-06	9	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH23705 +	374.00	255.72	6.8e-06	2	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH23705 +	372.00	264.07	2.3e-06	7	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AAH23794 -	367.00	259.36	4.2e-06	89	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH23688 +	367.00	259.26	4.3e-06	9	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH88313 -	366.00	253.98	8.5e-06	1	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH88313 -	366.00	245.87	2.4e-05	5	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH8316 -	366.00	245.87	2.4e-05	5	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH21501 -	366.00	242.37	3.7e-05	80	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH52942 +	362.50	262.01	3.0e-06	4	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AB123160 +	358.00	250.93	1.2e-05	1	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH09469 -	357.50	239.59	5.4e-05	50	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AB123161 +	357.00	252.68	1.0e-05	8	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAH3556 +	355.00	247.23	2.0e-05	14	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAH05181 +	355.00	247.23	2.0e-05	14	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV20477 +	354.00	244.62	1.0e-05	14	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AB107021 +	354.00	244.62	1.9e-05	9	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AB107020 +	354.00	244.38	1.9e-05	1	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AB106949 +	348.50	245.31	2.6e-05	9	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAH35491 -	346.50	225.60	0.0003	11	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAH76999 +	344.00	240.53	4.7e-05	1	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH2301 +	343.00	232.87	0.0001	30	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AB105593 +	342.50	243.56	3.2e-05	6	
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AB106948 -	342.50	234.52	0.0001	2	

/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:AAQ45339 + 341.50 238.81 5.9e-05
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AB109562 + 340.00 242.61 3.6e-05
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AB106515 + 340.00 240.57 4.7e-05
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AB106514 + 340.00 238.82 5.9e-05
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH79277 + 340.00 230.70 0.0002
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA160201

seq_documentation_block:

ID AA160201 standard; cDNA: 4871 BP.
XX
AC AA160201;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4190.
XX
KW Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
XX 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HXSE-) HXSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao Q, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
XX
PT P-PSDB; AAM41045.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
PS Claim 1: SEQ ID NO 4190; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with noctropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 4871 BP; 1815 A; 1021 C; 990 G; 1045 T; 0 other;


```

1129 eValAlGlyLysProSerSerIleIleLysAsnValThrThrLysProSer 1145
|||||
3595 GTGTAGAGAAACCTGCTAGTATTATAAAAAATGTTAGTACAAAGCCATCA 3644
1146 AlATrAlAlaLysThrGluLysGluSerGluGlnProGluLysLeuG1 1162
|||||
3645 AATATAGTCAAGTATCCTGAGAAAGAAAGAGCCATCCAGAAATTCATCA 3694
1162 nLysLeuProLysGluAlaSerHisGluLeuMetGlnHisGluLeuArgS 1179
|||||
3695 GAAATTCACCAAGACGTGACCATGAATCATACATCAATGAGGTMTAAA 3744
1179 eSerLysGlySerAlaSerSerGluLysGlyAlaLysAspArgGlu 1195
|||||
3745 GTTCAAAAACCTGTCATCTAGTGAAGAAAGGAAACCAAGATCGAGAT 3794
1196 HisSerGlySerGluLysAspAsnProAspLysArgLysSerGlyAlaG1 1212
|||||
3795 TATTCAGTGTGGAAAGGAAACCTGAAAGGAAAGGAAAGACGACTCA 3844
1212 nProAspLysGlySerThrValAspArgLeuSerGluGlnGlyHisPheL 1229
|||||
3845 GCCAGAAAGAGAGTATTTGGACCGCTGGAATGAAACAAAGAAATTTTA 3894
1229 yStrLeuSerGlnSerSerLysGluThrArgThrSerGluLysHisGlu 1245
|||||
3895 AAAGTGTCTCATATCTTCCAAAGAGGCTAGAACGTCAGATTAACATGAT 3944
1246 SerValArgGlySerSerAsnLysAspPheThrProGlyThrAspLysLy 1262
|||||
3945 TCCAGTGTGCTTCTCAATTAAGACTTCACTCCCAATAGAGCAAAA 3994
1262 sValAspLysAspSerArgAspLysSerSerSerLysArgArgAspGluA 1279
|||||
3995 AACTGACTATGACACAGAGAGTATTCAGTCCAAAGCTAGAGATGAAA 4044
1279 rGgLyGluLeuAlaArgArgLysAspSerProProArgGlyLysGluSer 1295
|||||
4045 AGAATGGAATTAACAAAGAGAAAGAGCTCTCTCGAAATTAAGATTCT 4094
1296 LeuSerGlyGlnLysSerLysLeuArgGluGlnLysArgAspLeuProLysLy 1312
|||||
4095 GCATCTGACAGAAATTAACCAAGGAGAGAGATTTGCCCTAATAAAA 4144
1312 sGly...AlaGluSerLysLysSerAsnSerSerProProArgAspLysL 1328
|||||
4145 AGGAACAGAGAGATTCAGAAAGAAAGTAAATTCAGTCCCAAGACAGAAA 4194
1328 ySPronHisAspHisLysAlaLysProGlyGlnThrLysArgProGlyGlnGlu 1344
|||||
4195 AACCTCATGATCACAAAGCCACTTATGATTAACGGCCCAATGAAAGAG 4244
1345 ThrLysProValAspLysAsnSerGlyLysGluArgGluLysHisAlaAl 1361
|||||
4245 ACNAAATCTGAGATTAATAATCTTGTAGAGATGCTGAAAGCATGTTT 4294
1361 aGluAlaArgAsnGlnLysGluSerSerGlyAlaAsnGlyHisValTyr 1377
|||||
4295 AGAAGCAAGAAACAATAAAGAGTCAAGTGCATATAACTACTATTATATAC 4344
1378 LeuThrArgGlnThrLeuProTTPArgArgSerTrpLeuLeuGlyArgTr 1394
|||||
4345 TTAACCCACCGAGACACAGGTTGAAAGGAAAGCAATTAATCTGGGCAATTT 4394
1394 pArgArgAlaLysSerSerArgAsnProSer 1404
|||||
4395 GACAAGAGTACTGTCAAGCTTAACCCAGT 4425

```

seq_name: /SIDSL/gcgdata/geneseq/geneseq - emb1/NA2001A.DAT:AA158415

seq_documentation_block:
ID AA158415 standard; cDNA; 4215 BP.
XX

```

AC AA158415;
XX 22-OCT-2001 (first entry)
DT Human polynucleotide SEQ ID NO 618.
XX
DE
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
XX Homo sapiens.
OS
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX
XX 25-APR-2000; 2000US-0552317.
XX
XX 09-JUL-2000; 2000US-0598042.
XX
XX 19-JUL-2000; 2000US-0620312.
XX
XX 03-AUG-2000; 2000US-0653450.
XX
XX 14-SEP-2000; 2000US-0662191.
XX
XX 19-OCT-2000; 2000US-0693036.
XX
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Dimañac RT;
XX
XX WPI: 2001-442253/47.
XX
XX P-PsDB; AAM39259.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 618; 10078pp; English.
XX
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening and
XX CC assays for receptor activity, arthritis and inflammation, leukaemia and
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX
XX Sequence 4215 BP; 1523 A; 901 C; 855 G; 934 T; 2 other;
XX

```

alignment_scores:

Quality: 5381.00 Length: 1381
Ratio: 4.436 Gaps: 15
Percent Similarity: 87.835 Percent Identity: 77.190

alignment_block:

US-09-811-045A-1 x AA158415
Align seg 1/1 to: AA158415 from: 1 to: 4215

```

1 MetMetGluValLysAspProAsnMetLysGlyAlaMetLeuThrAsnTh 17
  |||
158 ATGATGGAAGTGAAGATCTTAATATGAAAGGCGCAATGCTTACCAAC 207
  |||
17 rGlyLysTyrAlaIleProThrIleAspAlaGluAlaTyrAlaIleGly 34
  |||
208 TGGAAATATGCAATACCACTATAGATGACAGACATATGCAATTGGGA 257
  |||
34 yLysGluLysProProPheLeuProGluGluProSerSerSerGlu 50
  |||
258 ACAAGAGAACCTCCCTCTTACAGAGAGCCATCTCTCCAGAA 307
  |||
51 GluAspAspProIleProAlaGluLeuLeuCysLeuIleCysLysAsp 67
  |||
308 GAAGATGATCCTATCCAGATGATATGTTGTTCTCTATCTGCAAGATAT 357
  |||
67 eMetThrAspAlaValIleProCysCysGlyAsnSerSerCysAsp 84
  |||
358 TATGACTGATGCTGTGTGATTCCTGCTGTGGAAACAGTTACTGTGATG 407
  |||
84 lCysIleArgThrThrLeuLeuGluSerAspLysHisThrCysProThr 100
  |||
408 AATGTTAAGAACAGCAGCTCTGGAATCAGATGACACACATGTCGACG 457
  |||
101 CysHisGlnAsnAspValSerProAspAlaLeuIleAlaAsnLysPhe 117
  |||
458 TGCATCAAAATGATGTTCTCTGATGCTTATGTCCAATTAATTTT 507
  |||
117 uArgGlnAlaValAsnAsnPheLysAsnGluThrGlyTyrThrLysArg 134
  |||
508 AGCAGAGCGCTGAATTAATCTCAAAATGAACTGCTATACAAAAGAC 557
  |||
134 euArgLysGlnLeuProProPheLeuPheLeuValProProAspPro 150
  |||
558 TACGAAACAGTTACTCTCCACACACCCCAATACCACTCCGAGACCA 607
  |||
151 LeuSerGlnArgAsnLeuGlnProArgSerArgSerProIleLeuArg 167
  |||
608 CTGATTCACAGGACCTACACCTCTGATGAGATCTCCGATATCAAGAC 657
  |||
167 nGlnAspProValValPheArgTyrThrValSerProThrCysSerAsp 184
  |||
658 ACAAGATCCTCTATGATTCAGATGACATCTTCATCACT ..... 697
  |||
184 hLysThrAlaGlySerCysSerAspSerGlyThrLeuSerArgLeuPro 200
  |||
698 .....CACCA 703
  |||
201 AlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAlaPro 217
  |||
704 GCTCCGCTATATCTTCATTAATCTTAATCAAGTCTTCCCTGCCCCCTC 753
  |||
217 oValSerGlyAsnProSerSerAlaProAlaProAlaProAlaIleThr 234
  |||
754 TGGTCTGGAATCCGCTTCCTGCTCCAGCTCTGATCCGATATAACG 803
  |||
234 lathValSerIleSerValHisSerGluLysSerAspGlyProPheArg 250
  |||
804 CAACAGTATCCATATCAGTTCATGAGAAACATGATGACCTTTCCG 853
  |||
251 AspSerAspAsnLysLeuLeuProAlaAlaAlaLeuThrSerGluHis 267
  |||
854 GATTCGATATTAATATATTCGACGCTGACGCTTGCATCAGACACTC 903
  |||
267 rLysGlyAlaSerSerIleAlaIleThrAlaLeuMetGluGluLysGly 284
  |||
904 AAAGGAACTCCTCAATTCATACCGCTTATGGAAGAGAGGTT 952
  |||
284 aLProGlyThrSerProTyrPasnSerIlePheValGlyGlnSerLeu 300
  |||
953 A.CCAGGTGCTGTTCTTGGAAACCCATCTTGGTGGACATCAATATT 1002

```

```

300 uHisGlyGlnLeuIleProThrThrGlyProValArgIleAsnAlaAla 317
  |||
1003 GCATGACAGTGTATGCCCAACTGCTCCAGTGAAGAAATATACGCTGC 1052
  |||
317 rGProGlyGlyArgProGlyTyrPGLuHisSerAsnLysLeuGlyTyr 333
  |||
1053 GTCCAGGTGGTGGTCCAGCGCTGGGAACATTCCAACAACTGGCTAT 1102
  |||
334 LeuValSerProProGlnGlnIleArgArgGlyGluArgSerCysTyr 350
  |||
1103 CTGGTTCTCCACCACACAAATTTAGAGGGGAGAGAGCTGTACAG 1152
  |||
350 gSerIleAsnArgGlyArgHisHisSerGluArgSerGlnArgThrGln 367
  |||
1153 AATATTAACCGTGGGCGACACACAGCGAAGATACAGAGACTCAAG 1202
  |||
367 eProSerLeuProAlaThrProCysPheValProValProProPro 383
  |||
1203 GCCGCTACTACAGCACTCCAGTCTTGTACTGTCACACACCTCCT 1252
  |||
384 LeuTyrProProProProHisThrLeuProLeuProProGlyAlaPro 400
  |||
1253 TTGTATCCGCTCTCCCATACACTCTCTCCCGGGTGTCTCTCC 1302
  |||
400 oProGlnPheSerProGlnPheProSerSerGlnProProThrAlaGly 417
  |||
1303 TCCACAGTTTCTCTCAGTTTCTCTCTGCGCACGCCACCGCTGGGT 1352
  |||
417 ySerValProProProGlyPheProProAlaProAlaAsnIleSerThr 433
  |||
1353 ATAGTCTCCCTCCAGAGTTTCTCCAGCTCTCCCATTTATATCAAC 1402
  |||
434 AlaCysPheSerProGlyValProThrAlaHisSerAsnThrMetPro 450
  |||
1403 CCTGGGTATCATCAGAGAGTGCAGACGCTCATTAATAATCCATCC 1452
  |||
450 rThrGlnAlaProLeuLeuSerArgGluPheTyrArgGluGln .... 465
  |||
1453 AACCAAGCACCACTTTGTCCAGGGAAGATTCATAGAGACACGAC 1502
  |||
466 .....Asn 466
  |||
1503 GACTAAAGAGAGAAAAAGTCCAGCTAGATGATTACAAAT 1552
  |||
467 AspLysGlyArgGlu ..... 471
  |||
1553 GATTTTGTAAAGAAATTGATGAATACAAAAGATTCAAAAGAGCGTGA 1602
  |||
472 .....SerLysPheProTyrSerGlySerSerTyrSer 483
  |||
1603 GCGCTATTTTCCAGGCTTAATCTCCCTATAGTGTCTCTGATATTC 1652
  |||
483 rGSerSerTyrThrAspSerSerGlnGlyLeuAlaGlnHisIleHisAl 499
  |||
1653 GAAGTTCATATCTTAATTAATCAAGATCTGGGCAACACGCTGACG 1702
  |||
499 aLeuThrLeuSerProSerAlaAlaHisThrLeuAspLeuHisAsp 516
  |||
1703 TCTTATCTCGATCATTCAGCCGCTCACAATTCCTGCTCTATTCAG 1752
  |||
516 lAspHisProProGluGluAlaGluAlaArgSerAlaMetIleValHis 532
  |||
1753 ACCTCCATACCAGAGAGGAGAGCAAGAGCCCAATTACCGTTAC 1802
  |||
533 MetProAspLeuMetAspIleAlaHisAlaArgSerArgSerProTyr 549
  |||
1803 GGTCTAGATCTCATGATATCATCGATCTAGGTCAAAGTCAACCCCT 1851
  |||
549 rArgArgTyrArgSerArgSerArgSerProProGluPheArgGlyGln 566
  |||
1852 CAGAGCTATCATTCACGATCAAGATCTCTCAAGCGTTAAGGACAGT 1901
  |||
566 eProThrLysArgAsnValProArgGluGluLysGluArgGlyTyrPhe 582

```



```

1181 LysolysSerAlaSerSerGluLysGlyArgAlaLysAspArgLysHse 1197
3740 AAAAACTGCTAGTAGAAAAAGGAAAAACCAAAATGCAATTAATTC 3789
1197 rGlySerGluLysAspAsnProAspLysArgLysSerGlyAlaGlnPro 1214
3790 AGTGTGGAAAAAGAGAAACCTTGAAAAAGAGAAACAGCAGCTCAGCCAG 3839
1214 sPLysGluSerThrValAspArgLeuSerGluGlnGlyHisPheLysThr 1230
3840 AGAAAGAGAGTAATTTGGACCGCTGCAATGACAAGAAATTTAAAAAGT 3889
1231 LeuSerGlnSerSerLysGluThrArgThrSerGluLysHisGluSerV 1247
3890 CTGTCTCAATCTCCAAAGAGGCTAAGAACCGTCAGATTAACATGATTCCA 3939
1247 aLArgGlySerSerAsnLysAspPheThrProGlyArgAspLysVal 1263
3940 CTCGCGCTCTCCTCAATTAAGACTTCACTCCCATTAAGACAAAAAAACT 3989
1264 AspLysAspSerArgAspLysSerSerSerLysArgArgAspGluArg 1280
3990 GACTATGACACACAGAGAGATTAATCAAGTTCCAAAGTGAATGAAGAA 4039
1280 yLLeuLysLysArgLysAspSerProProArgGly..... 1292
4040 TGAATTAACAAGACGAAAAAGACTCTCCCTCTCGGGAATTAAGATTCTGG 4089
1293 .....LysGluSerLeuSerGlyGlnLysSerLysLeu... 1303
4090 CATCTGGACCGACGAAAAATAATACCAAGGAGGAAGAGAGATTGNC 4139
1304 ..ArgGluGluArgAspLeuProLysLysGlyAlaGluSerLysSe 1319
4140 TAAAAAAGGACCGGGGNTTCCCAAAAAAAGCT..... 4172
1319 rAsnSer...SerProProArgAspLysLysPro 1329
4173 .AATTCCTAGGCTCCCGGAGGGGGCGGAAACT 4204

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAH31144
seq_documentation_block:
ID AAH31144 standard; cDNA; 3529 BP.
XX
AC AAH31144;
XX
DT 27-JUL-2001 (first entry)
XX
DE Human colon cancer cell line KML2L4-A cDNA library derived seq#1078.
XX
KW Human: diagnosis; colon cancer; cancer; malignant; chromosome mapping;
XX
RW detection; colon cancer cell line KML2L4-A; ss.
XX
OS Homo sapiens.
XX
PN WO200018916-A2.
XX
PD 06-APR-2000.
XX
PF 23-SEP-1999; 99WO-US22226.
XX
PR 28-SEP-1998; 98US-0102161.
PR 28-SEP-1998; 98US-0102180.
PR 29-SEP-1998; 98US-0102280.
PR 08-OCT-1998; 98US-0103815.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR ) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;

```

PI Reinhardt C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamsom G, Drmanac R, Crkvenjickov R, Litwson M, Drmanac S, Labat I;
PI Lesnikowitz D, Kita D, Garcia V, Jones JW, Stachle-Crain B;
XX WPI; 2000-293155/25.

DR
XX
XX Polynucleotide library comprising 1079 defined sequences, useful in
PT the form of an array to detect cancer or susceptibility to cancer -
PS
PS Claim 1; page 501-502; 502pp; English.

CC The present invention describes a library of polynucleotides comprising
CC 1079 nucleotide sequences (given in AAH31145 to AAH31145). Also described
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
CC one of the 1079 sequences; (2) a recombinant host cell containing (1);
CC (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that
CC specifically binds to (II); (5) a vector comprising (I); and (6) a method
CC of detecting differentially expressed genes correlated with a cancerous
CC state of a mammalian cell comprising detecting a gene product encoded by
CC 65 of the 1079 sequences given in the specification. The polynucleotides
CC are used to monitor patients having (or susceptible) to cancer to detect
CC potentially malignant events at a molecular level before they are
CC detectable at a gross morphological level. The polynucleotides are also
CC useful for monitoring the efficacy of various therapies and preventive
CC interventions. Polynucleotide probes based on the disclosed sequences
CC are useful for chromosome mapping and detection of transcription levels.
CC The 1079 polynucleotide sequences were derived from a human colon cancer
CC cell line Km12L4-A cDNA library.
XX
XX Sequence 3529 BP; 1248 A; 778 C; 692 G; 803 T; 8 other:

alignment_scores:
Ratio: 4.251.00 Length: 1074
Percent Similarity: 87.337 Percent Identity: 78.492

alignment_block:
US-09-811-045A-1 x AAH31144 ..

Align seg 1/1 to: AAH31144 from: 1 to: 3529

1 Metctcgluvallylsasproasmetllysllylametleuthrantsrh 17
|||||
378 ATGATGGAACTGAAGAGATCCTAAATATGAAGAAGTCAATGTTACCAACAC 427
|||||
17 rglvlystyralleprthrllaeapalaglualatyralalegilyl 34
|||||
428 TGGAATAATATGCANTACCACACTATAGATGCCAAGCATATGCATTGGGA 477
|||||
34 ysiysgiulysprophelendprogugluiproserserserglu 50
|||||
478 AGAAGAGAAACCTCCCTTTACACAGAGGCCATCTTCCTCCCAAAA 527
|||||
51 Gluaspsaproilieproalagluendensylseuilecyslysaspil 67
|||||
528 GAACATATCTCTATCCCGAGATGAATTGTGTGTCTCATCTGCAGAGATAT 577
|||||
67 emetthraspalavalallieprocyscysglylasnsersercysasp 84
|||||
578 TATGACATGATGCTGTGTGATCCCTGCTGTGGAACAACGTACTGTGATG 627
|||||
84 lucyiliatrgtthrthleuenciluserasplyshisthrcysprothr 100
|||||
628 AATGTATAAGAACAGCACTCTGGAATCAGATGAGCACACATGTCGACG 677
|||||
678 TGTGATCAAATATGATGTTTCTCGATGCTTTAATTGCCAAATAATTTTT 727
|||||
101 CyshtislnasnaspvalserProAsPalaleutlealaasnlysPhele 117
|||||
117 uatginalavalasnasnphelysasngluthrglytyrrthrlysatgl 134
|||||
728 ACAGCAGGCGTATAATTAATTCAAAAAAGAAACGCGCTATCAAAAAAAGAC 777

134 euArglysglnLeuProPheLeuPheValProProProArgPro 150
|||||
778 TACGAAACAGTTACCTCCTCCACACCCCAATACACCTCCGAGACA 827
131 LeuSerGlnArgAsnLeuGlnProArgSerProIleLeuArgI 167
|||
828 CTGATTTCAGAGAACCTACACCTCTGATGAGATCTCGATTTCAGACA 877
167 nGlnAspProValIlePheArgTyrThrValSerProThrCysSerAspT 184
|||||
878 ACAAGATCTCTTATGATTCCAGTACATCTTACCACT..... 917
184 hrLysThrAlaGlySerCysSerAspSerGlyThrLeuSerArgLeuPro 200
918CACCA 923
201 AlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAlaProPr 217
|||||
924 GCTCGTATAATCTTCATTAACTTCTAATCAGTCTCTTGCCCTCC 973
217 oValSerGlyAsnProSerSerAlaProAlaProAlaProAlaIleThra 234
|||||
974 TGTGTTCTGAAATCCGCTCTTCTGCTCCAGCTCTGTACCTGATTAACG 1023
234 IarThrValSerIleSerValHisSerGlyLysSerAspGlyProPheArg 250
1024 CACAGTATTCATATCAGTTCATTCAGAAAATCAGATGGACCTTTGCG 1073
251 AspSerAspAsnLysLeuLeuProAlaAlaIleuThrSerGlnHisSe 267
1074 GATTCTGATAATAAATATGTCACCTCAGCTCTTGATGACAGCACCTC 1123
267 rLysGlyAlaSerSerIleAlaIleThrAlaLeuMetGluGlyLysGly 284
|||||
1124 AAGGGAACCTCTCAATTGCAATTACCGCTCTATGAGAGAAAGGGT 1172
284 alProGlyThrSerProThrAsnSerIlePhe_ValGlyGlnSerLeu 300
|||||
1173 A.CCAGGTGCGCTGTTGGAAACCCCATCTGTTGGACAGTATTATT 1222
300 uHisGlyLysLeuIleProThrThrGlyProValArgIleAsnAlaIa 317
|||||
1223 GCATGACAGTTGATCCACACACGTCTCAGTAAGAAATTAATCTGCTC 1272
317 rGProGlyGlyLysProGlyTyrGluHisSerAsnLysLeuGlyTyr 333
|||||
1273 GTCCAGGTGCTGTCAGCAGCTGGAAACATTCACACAACTTGCTAT 1322
334 LeuValSerProProGlnGlnIleArgArgGlyLysArgSerCysTyrAr 350
1323 CTGGTTCTCCACACCAACAAATTGAGAGGGGAGAGAGAGCTGTACAG 1372
350 gSerIleAsnArgLysArgHisHisSerGlyLysArgSerGlnArgThrGlns 367
1373 AAGTATAAACCGTGGCGACACACAGCGAAAGATCACAGAGAGCTCAAG 1422
367 eProSerLeuProAlaThrProCysPheValProValProProProPro 383
|||||
1423 GCCCTCACTACACAGCACTCAGTCTTTGTAACCTGTTCCACACCTCTCT 1472
384 LeuTyrProProProHisThrLeuProLeuProProGlyValProPr 400
1473 TTGTATCCGCTCCCTCCCATACACTCTCTCCCTCCGGGTGTTCTCC 1522
400 oProGlnPheSerProGlnPheProSerSerGlnProProThrAlaGlyT 417
1523 TTCACAGTTTCTCTCTCAGTTCTCTCGGACGACCCACCCCTGGGT 1572
417 yrSerValProProProGlyPheProProAlaProAlaAsnIleSerThr 433
1573 ATAGTCTCTCTCTCCAGGTTTCTCTCAGCTCTCGCAATTATACACA 1622

434 AlaCysPheSerProGlyValProThrAlaHisSerAsnThrMetProTh 450
1623 CTTGCGGATATCATCAGAGTCAGACAGCTCATTTCAAAATACATCCAAC 1672
450 rThrGlnAlaProLeuLeuSerArgGluGluPheTyrArgGluIn.... 465
1673 AACACAAACACCACTTGTGTCAGGAAAGAAATTTATAGAGACGAGCAG 1722
466Asn 466
1723 GACTAAAGAGAGAGAAAAGAAAAGTCCAAAGCTAGATGAGTTTCAAT 1772
467 AspLysGlyArgGlu..... 471
1773 GATTTTGCTAAGAAATTGATGAAATACAAAAGATTCAAAAGACCGTAG 1822
472SerLysPheProTyrSerGlySerSerTyrSera 483
1823 GCGTCATTTTCCAGGTCTAAATCTCCCTATATGTCGTTCTGTAATCA 1872
483 rGSerSerTyrThrAspSer_SerGlnGlyLeuAlaGlnHisIleHisAl 499
1873 GAAGTTCATATATCTTATTTAATCAAGATCTGGGTCAACAGCTCAGCG 1922
499 alLeuThrLeuSerProSerAlaAlaHisThrLeuAspLeuHisAspH 516
1923 TCTTATCTCGATTCATTCAGCCGCTCACATTCCTCTCTATTCAGGTC 1972
516 isProHisProProGluGluAlaGluAlaArgSerAlaMetIleValHis 532
1973 ACCTTCATACCCCAAGAGAGGAGAGAGAGCCGCAATTAACCGTTAC 2022
533 MetProAspLeuMetAspIleAlaHisAlaArgSerArgSerProTyr 549
2023 GTCTA.GATTCATGATATTCAT.CGATCTAGGTAAAGTCAACCCCTTA 2070
549 rArgArgTyrArgSerArgSerArgSerProProGluPheArgGlyGlns 566
2071 CAGACGTATCATTCACGATCAAGATCTCTCAACCGTTTAGGAGACAGT 2120
566 eProThrLysArgAsnValProArgGluGluGlyLysGluArgGluTyrPhe 582
2121 CTCTTAATAAACGTAAATGTACTTCAAGGGGAAAGCAACGTAAATTTT 2170
583 AsnArgTyrArgGluValProProProTyrAspIleLysAlaTyrTyrG 599
2171 AATAGATACAGAGAAATTCACACACATATGACATGAAGACATATTAGG 2220
599 yArgSerValAspPheArgAspProPheGluLysGluArgTyrArgGluT 616
2221 GAGAAAGTGTACTTAAAGACCCATTGGAAGAAAGAACGTAACCGAAT 2270
616 rPGLuArgLysTyrArgGluTyrPyrGlyLysTyrTyrLysGlyTyrAla 632
2271 GGGAGAGAAATATAGAGAGTGTATGAAATATTTATTAAGCTTATCT 2320
633 ValGlyAlaGlnProArgProSerAlaAsnArgGluAspPheSerProG 649
2321 GCTGAGACACASCCATAGACCTCAGCAATATAGAGAACTTTTCCAGA 2370
649 uArgLeuLeuProLeuAsnIleArgAsnSerProPheThrArgGlyArga 666
2371 GAGATTTTGGCCACTTAACTCMGAAATTCCTCTCACAAGAGGCCCA 2420
666 rGGLuAspTyrAlaAlaGlyLysSerHisArgAsnArgLysGly 682
2421 GAGAGACTATGTTGTGGCAAAAGTCTATGAAAGTCGAAATAGTAGC 2470
683 AsnTyrProGlyLysLeuSerThrArgAspSerHisAsnAlaLysAspAs 699
2471 AACTATCCAGAAAAGCTTTCAGCAAGATGTCTCAATTCAGAAAGATAA 2520
699 nProLysSerLysGluLysGluSerGluAsnValProGlyAspGlyLysG 716


```

1  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2521 TCAAAAGTCAAAAGAGAGAGAGTGAACCGCTCCAGAGAGATGATGAAG 2570
716 lYsAluSlySHslySlySHsArGLysArGLysngluLysglYglu 732
|||||
2571 GAAATAGCATAGTAAACACAGAAAAAGAGA.....AAAGGGAG 2611
733 GluSerGluSerPheLeuAsnProGluLeuGluThrSerArgLyscy 749
|||||
2612 GAAAGTAGAGGTTTCTGAACCCAGATATATAGAGACTTCTAGGAATC 2661
749 sArGLysSerSerGlyLeaspGLuThrLysThrAspThrLeuPheVal 766
|||||
2662 AAGAGAACTACAGCTGTGAAAGAAATAAACAAGACTACTTTTGTTC 2711
766 euProSerArGLysAspAlaThrProValArGLyspGLuProMetaSpAla 782
|||||
2712 TCCCAAGTAGAGATGATGCCACACTGTATAGAGATGACCAATGATGCA 2761
783 GluSerIleThrPheLysSerValSerAspLysAspLysArGLysAs 799
|||||
2762 GAATCATACCTTTAATCATGCTGTGAAAAAGACAAGAGAGAAAGGA 2811
799 pLysProLysValLysSerAspLysThrLysArGLysSerAspGLysSerA 816
|||||
2812 TAAACCAAAAGCAAGGTGATTAACCAACGGAAGAGATGATCTG 2861
816 lAThrAlaLysLysAspAsnValLeuLysProSerLysglYProGlu 832
|||||
2862 CTGTCTCCAAAAAGAAATATGTAAACCTGTAAAGAGACCCCAAGAA 2911
833 LysValAspGLysAspArGLysSerProArGLysSerGluProLeu 849
|||||
2912 AAGGTAGTAGAGACGTGAGAGATCTCCGATCTGAACCTCAATTA 2961
849 sLysAlaLysGLuGluAlaThrLysIleAspSerValLysProSerSerS 866
|||||
2962 AAAGGCCAAAGAGAGACTCCGAAGACTGACATACTAATCATCATCT 3011
866 eArSerGluLysAspGluLysValThrGlyThrProArGLysAlaHisSer 882
|||||
3012 CCTCTCAGAGAGTGAATAAATCATCTGAACCCCAAGAAAGCTCATCT 3061
883 LysSerAlaLysAsp..ThrArGLysInSerGluProArGLysThrArGLys 899
|||||
3062 AAATCAGCAAAAGACACCAAGAAACAAACAGTCAAAAGAGCAAAAGT 3111
899 eArLysArGLuThrValProLysThrSerSerGluLysSerGluProValArGL 915
|||||
3112 GAAAGAGACTATTCCAAGATGTCAATCAGAAAAGCTAACAACTAAG 3161
916 ThrArGLysProArGLysSerLeu...ArgLysIleAsnThrLeuIleAla 931
|||||
3162 AAGAAAAAGCCCAAGAACCTTAATGAGAAAAACCAACCACTTGATTA 3211
931 gGluLysAsnGluArGLuLysArGLys..LysSerValAspLysAspPh 947
|||||
3212 GGAGAAAAAGAAAAAGAACTGAAGAAAAAGCGCTAGATAAGATTT 3261
947 eGluSerSerSerMetLysIleSerLysValGluGlyThrGluIleVal 964
|||||
3262 TGAGTCTTCTCAATGAAATCTCGAAACTAGAAAGAGACTGAATACTGA 3311
964 ySProSerProLysArGLysMetGluGlyAspValGluLysLeuGluArGL 980
|||||
3312 AACCATCACCACCAAGCCCAAAATGGAACCTGATCTGAAAAAAGGRTAG 3361
981 ThrProGluLysAspLysIleAla..SerSerThrThrProAlaLysLysI 997
|||||
3362 ACCCTGAAAAAGACAATTTCTTTAAGTGCACACC.....AAAAAA 3405
997 lElyLysAsnArGLuThrGlyLysLysIleGlyAsnAlaGluAsnAla 1013
|||||

```

```

3406 TCAACTCAACAGAGAACTGGGAGAGAAATTTGAGAGTACAGAAATATA 3455
1014 SerThrThyLysGluProSerGluLysLeuGluSerThrSerSerLysI 1030
|||||
3456 TCAAACCAAAAGAACCTCTGAAAAATTGAGTCAACATCTGCAAAAGT 3505
1030 eLysGluGluLys 1034
|||||
3506 TAAAAAAGAAAAA 3518
seq_name: /SID1/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT: AAS28668
seq_documentation_block:
ID AAS28668 standard; DNA; 17252 BP.
XX
AC AAS28668;
XX
DT 07-NOV-2001 (first entry)
XX
DE Genomic sequence #508 encoding for novel human respiratory antigen.
XX
KW Human; respiratory antigen; respiratory disorder; throat disorder;
KW lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;
KW anti allergic; anti asthmatic; anti inflammatory; olfactory;
KW respiratory active; ds.
XX
OS Homo sapiens.
XX
WO20015448-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01333.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0188874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226881.
PR 22-AUG-2000; 2000US-0226886.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.

```


119 1n..... 119
11 AGGTACCTGCTGATCATTTTATGAAAAAATCTTTTAAGTATTA 6151
119 119
6152 ACTGACTTCAGTGAATACGATTAACATTTTCTGATTAATATGCTTG 6201
120AlaValAsnAsnPhelysAsnGluThrIleGlyTyrThrLeuA 133
6202 GATATCTGATGGCTGTAATTAATTAAGTAACTGAACTGGCTATACAAAA 6251
133 rGleuArGlysglLeuProPheLeuPheValProProPArg 149
6252 GACTAGAAAACAGTTACTCTCCACACACCCCAATACCACCTCCGAGA 6301
150 ProLeuSerGlnArgAsnLeuGlnProArgSerArgSerProIleLeuAr 166
6302 CCATGATTCAGAGAACTACAACTCGATGATGATCTCCGATATCAAG 6351
166 gGlnGlnAProValAlaPheArgTyrThrValSerProThrCysSerA 183
6352 ACACACAGATCCTTATGATTCACAGTACATCTCATCAACT..... 6394
183 sPThrIysThrAlaGlySerCysSerAspSerGlyThrLeuSerArgLeu 199
6395CAC 6397
200 ProAlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAlaPr 216
6398 CCAGTCCGCTGATATCTTCAATTAATCTTAATCACTGCTTGGCCCC 6447
216 oProValSerGlyAsnProSerSerAlaProAlaProValProAspIleT 233
6448 TCCTGCTGTGGAAATCCGTCCTCTGCTCCAGCTCTGACCTGATATA 6497
233 hrAlaThrValSerIleSerValIleSerGluIubSerAspGlyPro... 248
6498 CTGCACACATATCCATATCAATTCAGTTCAGAAAAATCAGATGACCTTTT 6547
248 248
6548 CGGTAAAGCTGTGTGTTTCACTGTTAGAAACCAATGAGTCATGAT 6597
248 248
6598 GTAGTGTGTTGTGTGTTATCACTTTTAACAGCTTTCTTAGTGACCA 6647
248 248
6648 CTGTGCTCAGTAATGGGATGACTTGTGAAGAAGTACTAGTCTTAATG 6697
248 248
6698 GGGGAAGAGAAAAACAAGTATGATTTGGCAATGAGATTAAAGAT 6747
248 248
6748 CCTGAGGTCAAAATACTACAGAGTGCACACGCCCTATTAAAGTTG 6797
248 248
6798 AATGTTCCATATACCAACAGCTGAACTTAACAGAGGTGGAGAAAT 6847
248 248
6848 CACTTGATCATATGATGATTTTGGATPAGAGTAGTTTGAAGTAAC 6897
248 248
6898 TAAGAACAGTTTAGATAGTAGAGAGAGGGGTAAAAACACAGATAAGT 6947

248 248
6948 AGAGAAATCAGGAATATGATGATTTGTTGGACATAATATATATAGTAGAG 6997
248 248
6998 AATAGTAAAAACTCACAGCTTACCACTTCAGAAATGAGACTTTGTCCAT 7047
248 248
7048 GTAGTTAATGTCCTCAATATCTTTAAAAAAGAGGTGAAG 7097
248 248
7098 AGGCCATTTTCAGACACACTGAAGCAAAATCTTAGATGAGACCTTAC 7147
248 248
7148 ATTTACATTTAAGAAATTCACATGATCTGTGGAAATTAATTTGGTAGATA 7197
248 248
7198 GTGGAAATAGGCCAGAGAGTTCTATGCTGACCCAGATTAATCTTTCA 7247
248 248
7248 GTTTAGGTAGAACATAGACCTAAACTTAACCTTAATCCCTAGGGGAT 7297
248 248
7298 TTTTGTCTCATAGCCAGACACTTGTGATTTTAAGTTTACTATCTTGCAA 7347
248 248
7348 CTGATTTCTGTCAATACCTAGATCTCTTAAGAGGTCCTTCATTTTA 7397
248 248
7398 CTTTCCACTTACTACATATAGAGAACTCTCAAAAGTTTCAGGCTTCTC 7447
248 248
7448 CACTTCTCAGAAATTCATCATCTTTGACTTAACATCTTAAGTGC 7497
248 248
7498 AGGTAGTGGTAAACAGCAAACTTTTACTCTTAATTCATATTC 7547
249Phe..ArgAspSerAsp 254
7548 TGATTAATTTTAATGATTAATTAATTAATGTTTAAAGGATTTCTGATA 7597
254 snLysLeuLeuProAlaIleAlaLeuThrSerGluIubSerIysGlyAla 270
7598 ATAAATATATGCGACGCTGCTTGCATCAGACGACCAAGGAGACC 7647
271 SerSerIleAlaIleThrAlaLeuMetGluGlu..... 281
7648 TCCCTCAATTCCAATTAACCGCTTATGGAAGAGAAATTTTAAGTGC 7697
281 281
7698 TGCTTAATTTGGATTTTAAACAGAGGTTTGTCTTCTGCGCTT 7747
281 281
7748 GTTCTACCTCTTTTCTTTTCTTTTAACTAATTCCTGTTATTTGNG 7797
282LysGlyValProGlyThrSerProThrAsnSerIlePhe.V 295
7798 GTTGTCTTAAGGTTTACAGGAGGCTGTCTTGGAAACCCATCTTTGCG 7847
295 aLgIubInserLeuLeuHisGlyIubLeuIleProThrThr..... 308

```

7848 TTTGACAGTCATATTGTCATGAGACAGTTGATCCCACTGGTGGAGTAA 7897
308 ..... 308
7898 GATCAGCTTGGTTAGACCTTTTCCCTTAAAAAATTAATTAGCTGA 7947
309 .....G 309
7948 TTTAATGTGCATTTTGTATTACCTACGCTTTTATTTTTGTTCCTAG 7997
309 lypOvalaIagIleasnaIaIaIarProGIyGIyGIyArqProGIyTir 325
|||||..... 325
7998 GTCAGTAAGAAATAAATACACGCTGCTCCAGGTGGTGCACAGGCTGG 8047
326 GIu..... 326
8048 GAACAGTGAATGATGTTTACATAATCTTCAGATGATGCTGCCAAGCT 8097
326 ..... 326
8098 AGATAATGAAGGTCCAACTAGGCAAACTAGATAATGAAGGTCCAAC 8147
326 ..... 326
8148 TAGGCACGGGTGCTGCTCTTATTGTAACTTCATGTGCCAGATTAA 8197
326 ..... 326
8198 CATTTTTATGTAATAATGTCAAAATTCAGCTTATATGAGGTTTTTTT 8247
326 ..... 326
8248 AAATGACCAATCTTTCTAAATACAAAATTTTAAAGTTTTTTCTAC 8297
326 ..... 326
8298 TTTCAATCCACAATAAGGGAGCTTGACAGACTTCAGTTTAATGATA 8347
326 ..... 326
8348 GATAAGATTACTGAAAAGTTATTTAAATATGTGGCAGTTATTTTCC 8397
326 ..... 326
8398 TAGTAGTTTAACCTTTTCTACCTTAATTGATGATCAAAAGTTGATT 8447
326 ..... 326
8448 TGAGTACATAATACATPATATAGATTACGTCCTCCAAACAAAATTTT 8497
326 ..... 326
8498 TGCTGAATCGAGTTGTTCTTGCCACCTAAATACAAAGACTTTGAT 8547
326 ..... 326
8548 TTTCTGTGGGATTTTATGCAATAAAAATCAGCAGACTTACTTAAT 8597
326 ..... 326
8598 GTAATTTTAGTAAATATTAATTCTCTGTTATATATCCGATTGAAAAAT 8647
326 ..... 326
8648 GATATATTCTTAAAAAGTATGAGTTACTAGCAGATGTATCAGTAGAAGA 8697
326 ..... 326
8698 ACATAGATTAGAAACAGTAGACTTGGTTTGAATCCCAATCTCTTCA 8747
326 ..... 326

```

```

8748 TTTCTAGCTGCATGACAGTGGCTTGTGTTTACTCCCTAACATCCAT 8797
326 ..... 326
8798 TTCCTGTTTGATTTGGAATGACACTACCTTAACATGTAAATGA 8847
326 ..... 326
8848 CTTCAATTGTGCCGACGACAAAAGGCTCTTTCTTGTATATTAACAAT 8897
326 ..... 326
8898 AAGTTCACAGATTAATTATAGAGTGTATTAATAATATATTAATTA 8947
326 ..... 326
8948 AAGTAATCTATTTTGAAGCTAGCTATTGTATGATACAGTTAATGGTG 8997
326 ..... 326
8998 GGGGAAAGTTAGTTTATAGTAATAATATATTTACTCAGCTGTATCCG 9047
326 ..... 326
9048 AAACAATTTAAGACCTGGGAGTAGGAAAAAGTTTCAGAGTAGTTAA 9097
326 ..... 326
9098 ATTGCTCTTAACAGCTATGATGAGAAAGTCATGTTATTAACCATGG 9147
327 .....His.SerAsnIyLeu 331
9148 AGAAATCTTGAGCCTATTTTCTTCTCCACACAGATTCACAAACTT 9197
332 GIyTyrLeuValSerProProGIInIleArqArgIyGIuArgSerCy 348
|||||..... 348
9198 GGCATCTGTTCTCCACACAACAATAAGAGGAGGAGAGACTG 9247
348 sTyT..... 349
9248 CTACAGTAGGCAATGTAAAAATCTGAAAATATATCATCTTATTTT 9297
349 ..... 349
9298 TATATATATCTCACCAAGAATAATTAAGCACAGCAGTGGCAAGAAG 9347
349 ..... 349
9348 GGGGTCAATTGTTGTATTGTGTTATCCCTTAAGTCCCTGGGTAGT 9397
349 ..... 349
9398 TCCTTAAAGAGGCTAGAAAAAATAATATATATATATCTTAAATGA 9447
349 ..... 349
9448 TGTTTTTAAGACATTTTAAATATTCCTCCAGTGAATTACTTCTCCT 9497
349 ..... 349
9498 CTCAGCAACAGCAGAGATTTGGCTGCAATTTACCTCTTGAAGTAGAA 9547
349 ..... 349
9548 GAAAGTCAGGGGTGACACAATATCTAAGATACAGTTTAATTAAGATTA 9597
349 ..... 349
9598 AAGTTAACAGCTCAGTCCCACTGTGAACAGAAATTCACATATTTCT 9647
349 ..... 349
9648 CTCTAGACCCCACTGTGCCAATCATCTCCACGCTCTTTTCCAAAG 9697

```

349 349
9698 AAGACAGAGATGTTAGAGCTCCGCCCTAGTGGGGGTACAAAC 9747
349 349
9748 AATATTAACTAAGAATAGAACTCTACTGCTAGTCAAGGGTCCCA 9797
349 349
9798 ACCCTGGGCCACAGACAGATATCATCTTGCCCTGTAGAACTGGCA 9847
349 349
9848 ACACAGAGAGGTGACAGACAGCTGAGGAGACATTACCACTGAGCTC 9897
349 349
9898 ACCTCTTGATGATCAGACAGTGGCATTAAGATTCTCATAGAGACATGAAC 9947
349 349
9948 CCTGTGTGAAGCTGGGTGCAAGAGATCTGGTTGCACGTTCTTAGGA 9997
349 349
9998 GAATTAATGCTGATCATCTAAGGTGGAAGATTTTACCCCTGCCC 10047
349 349
10048 CCGTCATGAAAAAATGTCTTCCACAAAAACAGCTCCGTGTCCAAAAC 10097
349 349
10098 GGTGAGAACTCTAGCTAGTCAATATACACCAGATAAGAGACC 10147
349 349
10148 TCCAAAACTTCAACAGAGTTTGGCCAGCATATGCTCATGCTGG 10197
349 349
10198 TAATTTAGAGCTTTGGAGGCTGATGGGGAGATTGCTTGAGGCTGA 10247
349 349
10248 AGTTCAAGATTAGCCTGGCAACATAGCAAGACTGTCTTACCAAAAA 10297
349 349
10298 AATAAAAATACAGACATGGTGGCATGCACCTGTGCTCCAGCTACCTGG 10347
349 349
10348 GAGGTGAGGACAGAGATCATCTTGTAGAGTCAAGGCTGCAATG 10397
349 349
10398 AGCTAATACCATCATCTGCACTTAGCCTAGGGAGACAGATGAGACCT 10447
349 349
10448 GTCTCCAAAAAAGAAATGGGATATCTTCTAGTGTCTCATTTGCTAA 10497
349 349
10498 ACTTAATGAAGTTTACTATGCTATTACTGATCACAATTTAAAGAAGT 10547
349 349
10548 AATAGTTATTCATGGCTCACTGAAAGATGCTTAAATCTAAACCCCTAC 10597

349 349
10598 TGTGCTGTGTGTAACAATATTTATATTTAAAGCACCTCTAGTTAA 10647
349 349
10648 CAGAAGTAACAGATTCTGTTGGTGAGCAAAATAGAACTTAGCCAA 10697
349 349
10698 CACTGTTATTTTACTTAAGCTTTTGAGAGAAAGCGTATATTTTCATT 10747
349 349
10748 GTTAATATAGAAGTTAAACTAGATTTATCTGCTTTGTATGATCTTT 10797
349 349
10798 ACCCACTTCCTCCCAAAAAAGAGGTTACAGAGTTGGTTTCTTCATT 10847
349 349
10848 TTAAACTTTTGGGGCCAGGTGTGTGCTCTCATCTTAACCCAGCACT 10897
349 349
10898 TTGTGATATCAGCTGAGGTGAGAGTTTGAGAACAGCCTGCCAACATG 10947
349 349
10948 GTGAACCTGTCTCTCTAATAAAATACAAAAATTAAGTGGCATGGTGC 10997
349 349
10998 ACAGCTCTAGTCCAGCTACTTGGGAAGCTGAGGAGAGAAATTCCT 11047
349 349
11048 TGAACCCAGAGGTGAGGTTCAGTGAAGCAAGATTACACCACTGACT 11097
349 349
11098 CCAGCTGACAGACAGAGTGAAGACTGTCTCAAAAAATAAAAATAAT 11147
349 349
11148 GAATAAGTAAGCTTTGGGAGAGCTAGGTCAATAGCAGATTAACTGTC 11197
349 349
11198 AGTGCTTCAAAAAGACAGTTTAGAAGTAAACAGTGTATTAAGTGAGT 11247
349 349
11248 GATAATCACTGTAAGCTTCATAAAGAGAACTTTAAATACATATTTCC 11297
349 349
11298 GTGTAGAGAGGTCTATAACTTGATCATCTCAGTCGAGACATGGCT 11347
349 349
11348 TAAAAATTAATGTTGAGATTATTTTACAAATAAATGTACTAATGTG 11397
349 349
11398 CTACACTAGTGAATAGTTGGTTCTATTCACAGTGTTTAAAGTTGAAAA 11447
349 349
11448 TAATTAATGATGAGAGATAAGCTTACTTGTCTCAGATGACTTTAGTTT 11497
350 ArgSerIleasnArgIya 356

```
11498 GAAGAGTAATATCTTGGAAATTTATTTTCCAGAGTATAAACCGTGGC 11547
356 rghIshISerGluArgSerGlnArgThrlInSerProSerLeuProAla 372
11548 GACACACAGCGAAGAGATCAGAGAGCTCAAGGCCCTCATTACACAGA 11597
373 ThrProCysPheValProValProProProLeuTyProProProPr 389
11598 ACTCCAGCTTTGTACTGTTCACACACCTCTGTGTATCCGCCCTCC 11647
389 ohISThrLeuProLeuProProGlyValProProGlnPheSerProG 406
11648 CCATACACTTCCTCTCCCTCGGGGTTCCTCCACAGTTCTCTCTC 11697
406 lnpheProSerSerGlnProProThrlAlaGlyTyrSerValProPro 422
11698 AGTTTCCCTCCGCGCAGCCACCCCGCTGGGTATAGTCTCTCTCA 11747
423 .GlyPheProAlaProAlaAsnIleSerThrlAlaCysPheSerProG 439
11748 GGGGTTCCTCCAGCTCTGCCAATTTATCAACACCTCG .GTATCATCAG 11796
439 lYvalProThrlAlaHisSerAsnThrMetProThrlThrlAlaProLeu 455
11797 GAGTGCAGACAGCTCATTCAAATACCATCCACAACAACAGCACACCT 11846
456 LeuSerArg .GluGluPheTyrArgGluGln..... 465
11847 TTGTCCAGGGAGAGATTTCTATAGAGACAGCAGACTAAAGAGAGT 11896
465 ..... 465
11897 ATGATCTAAATTTGAATTTATATACACTTTTTCATTTTCTGTATATA 11946
465 ..... 465
11947 ACATTAATAGGTGTCTGAAAAAGTTTCCAGTGTTCAGTTAGTATAGA 11996
465 ..... 465
11997 GAAATTCACCTAAATTTAAACCAAGCCATACAGCAACTTCTCAGGGA 12046
465 ..... 465
12047 TTCTAGTTTATAGTTTAAAGTGAAAACTAATGCAATTCATTCAAT 12096
465 ..... 465
12097 TCATTCATGATCTATATACATTTGTCTTGTAGTTTCTTAATGGTTTT 12146
465 ..... 465
12147 TGTTAATCTTACTCTTAACTTATTTACTAGGAAAAAGAAAGTCC 12196
465 ..... 465
12197 AAGCTAGATGAGTTTACAATGATTTTGTAAAGAAATGATGAAATACA 12246
466 ..AsnAspIysGly..... 469
12247 AAAGATTCAAAAGAGCGCTAGCGCTCATTTTCCAGTTAGTGTTTGCA 12296
469 ..... 469
12297 AGCCTTACAACAGAAATTAACAAGCGACTTTGAATATCCAGGATTAAC 12346
469 ..... 469
12347 TATGATATTAAGAACATTTGCACTTAGGAATTTTCTGATTCGGTCTAG 12396
469 ..... 469

12397 TGTATAGAAAGTGAAAGTACCTTATTTCCAAATATCTGTTCAATGAAT 12446
469 ..... 469
12447 AATGCTCAGTGAATTTTATTTGTAGAACTAACAGTTGTGGGATTA 12496
469 ..... 469
12497 AATTGCCCTGTCCCTTAAGTTATTCATAGTATCAATAATATGGAAAT 12546
469 ..... 469
12547 GTTTTGGAGAAAGAAATATAGGGAAATAGCCAAAAAACAGTTCCACGA 12596
469 ..... 469
12597 CCGTTTACATAAGGCTACAGCTTCTGTGGAACATTAACCTGCTTA 12646
469 ..... 469
12647 TTATGAACTAGTAGTAGAATACAGATCGTTTATGTATAGAAAAAGTGA 12696
469 ..... 469
12697 TTATTAATAATTCCTCGAAGTTTGTTTAAAGACTTATTTTATAGCCTT 12746
469 ..... 469
12747 ACCTTGAATATGTGGTTCTTTTAAACTAGGCTGTGGCCAGTATAGA 12796
469 ..... 469
12797 CAAAAACAGACTCCTACTTCATATTCAGATAGAAAGTAAATTTAAAA 12846
469 ..... 469
12847 TTAATGACATTTTGTTAATGTATAGTACATATATACACAGGTACA 12896
469 ..... 469
12897 TTAATGCACCTTACCTTAATGTTAAGACTTCACCTTTGTTTGTAAACA 12946
469 ..... 469
12947 GTGATTCCTCCGATTTGTACTTATTCATTCACAGTAGAGTCACTGT 12996
470 ..... Ar 470
12997 GCTCTTAAATCATCTAAGATGAATTTGATATTAAGATAGAGAACCG 13046
470 gGlu..... 471
13047 TGAGAAAGAAATGCTAGAATCTGAGAGTCTGTGTTTATTTGTATGAT 13096
472 ..... SerIysPhePro 475
13097 TTCTCAACTATCAACTATTTGTTTGTGTTTGTATAGGCTAATATCCCC 13146
476 TyrSerGlySerTyrSerArgSerArgSerTyrThrPheSer.SerGln 492
13147 TATAGTGTCTCTCGATTTCAAGAGTTCAATATCTTATTTTAATATCAG 13196
492 lYleuAlaGlnHisIleHisAlaLeuThrLeuSerProSerAlaAlaHis 508
13197 ATCTGGTTCAACAGCTTCACGCTCTTATTCGATCATTCAGCCGCTCAC 13246
509 ThrLeuAspLeuLeuHisAspHisProHisProGluGluAlaGluAl 525
13247 ATCTCGTTCTTATTCACGCTCACCTCCATACCCAGAGAGAGAGAGCC 13296
525 ArgSerAlaMetIleValHisMetProAspLeuMetAspIleAlaHisA 542
13297 AAGAGCCGCAATTAACCTTCACGGTCTAGATCTCATGATATCAT .CGAT 13345
```

542 IARSerArgSerProProTyrArgTyrArgSerArgSer 558
 ::
 13346 CTAGGTCAGAGCCCTTACAGAGCTATCATCAGATCAAGATCT 13395
 559 ProProGluPheArgGlyGlnSerProThrLysArgAsnValProArgG 575
 ::
 13396 CCTCAAGCCTTTAGGGGACAGTCTCTATTAACGTAAATGTACCTCAAG 13445
 575 uGluysGluArgGluTyrPheAsnArgTyrArgGluValProProT 592
 ::
 13446 GCAACAGACGCTGATATTTAAATGATACAGAGAACTCCACCCAT 13495
 592 yASpIleLysAlaTyrTyrGlyArgSerValAspPheArgAspPrope 608
 ::
 13496 ATGACATGAAAGCATATATGGAGAGAGTGTACTTTAGAGACCATTT 13545
 609 GluysGluArgTyrArgGluTyrGluArgLysTyrArgGluTyrP 625
 GAAAAAGAACGCTACCGAGATGGGAGAGAAATATAGAGAGGTATGA 13595
 625 uLysTyrTyrLysGlyTyrAlaValGluValProArgProSerAla 642
 ::
 13596 AAAATATTATAAGGTTATGCTGACACAGCCTAGACCTCAGACAA 13645
 642 snArgGluAspPheSerProGluArgLeuLeuProLeuAsnIleArg 658
 ::
 13646 ATAGAGAGACATTTCTCCAGAGAGATTTTGCCACTTAACATCAGAA 13695
 659 SerProPheThrArgGlyArgArgGluAspTyrAlaAlaGlyGlnSer 675
 TCCTCCTTCACAAGAGCCCGCAGAGAACTATGTTGGTGGGCAAGTCA 13745
 675 sArgAsnArgAsnLeuGlyGlyAsnTyrProGluLysLeuSerThrArg 692
 ::
 13746 TAGAACTCAAACTAGGTAGCAACTATCCAGAAAGCTTTCGCAAGAG 13795
 692 sPSeHisAsnAlaLysAspAsnProLysSerLysGluLysGluSerGlu 708
 ::
 13796 ATGCTACAAATCGAAGGATATATACAAAGTCANAAAGAGAGATGAA 13845
 709 AsnValProGlyAspGlyLysGlyAsnLysHisLysHisArgLysArg 725
 AACGCTCCAGAGAGATGTAAAGGAATATAGCATATAGAAACACGAAAG 13895
 725 gArgAsnGluGluLysGlyGluLysGluSerPheLeuAsnProGlu 742
 ::
 13896 AAGA.....AAGGGAGAGAAAGTAGGGTTTCTGAACCCAGAGT 13936
 742 euLeuGluThrSerArgLysCysArgGlySerSerGlyIleAspGluThr 758
 ::
 13937 TATTAGAGACTTCTAGGAAATCAAGAAACCTACAGCTGTGAGAAAT 13986
 759 LysThrAspThrLeuPheValLeuProSerArgAspAspAlaThrPro 775
 AAAAAAGACTCATTTGTTGTTCTCCCAAGTAGAGATGATGCCACACTGT 14036
 775 IArgAspGluProMetAspAlaGluSerIleThrPheLysSerValSer 792
 ::
 14037 TAGAGTGAACCATGATGCAGAAATCAATCACTTTAAATCAGTGTCTG 14086
 792 sPLysAspLysArgGluLysAspLysProLysValLysSerAspLysThr 808
 ::
 14087 AAAAAACAAAGAGAGAAAGGATAAACCAAAAGCAAAAGGCTGATAAAC 14136
 809 LysArgLysSerAspGlySerAlaThrAlaLysLysAspAsnValLeu 825
 ::
 14137 AAACGGAAGATGATGATGCTGCTGTCTCCAAAAAGAAATATATTGAAA 14186
 825 sProSerLysGlyProGluGluLysValAspGlyAspArgGluLysSer 842
 ::
 14187 ACCTGCTAAAGACCCCAAGAAAAAGTAGATGAGAACTGAGAGATCTC 14236

842 roArgSerGluProProLeuLysLysAlaLysGluLunAlaThrLysIle 858
 ::
 14237 CTGATCTGAACTCCATTTAAAAAGCCAAAGAGAGACTCCGAAGACT 14286
 859 AspSerValLysProSerSerSerSerGluLysAspGluLysValThr 875
 ::
 14287 GACAAATAGTAATCATCATCTCTCTCAGAAAGATGAGAAAAATACACTG 14336
 875 yThrProArgLysAlaHisSerLysSerAlaLysAsp.ThrArgGluGln 891
 ::
 14337 AACCCCAAGAAAAAGCTCCTTAATCACCAGCAAAAGAACCAAGAAACAA 14386
 892 SerGluProArgThrArgArgSerLysArgThrValProLysThrSer 908
 ::
 14387 AACCACTCAAGAGAGAAAAAGTAAGAGAGACTATTCCAAAGATGTCAAA 14436
 908 rGluLysSerGluProValArgThrArgArgProArgSerLeu...Arg 924
 ::
 14437 TCAGAAAGCTAACCACTAAGAGAAAGGCCAGAGAGCTATATGAGAA 14486
 924 yLleAsnTyrLeuIleAlaArgGluLysAsnGluArgGluLysArgLys 940
 ::
 14487 AACCAAAACCTTGATATATAGGAGAGAAAAAGAAAGAAAAAACTGAAG 14536
 941 ..LysSerValAspLysAspPheGluSerSerSerMetLysIleSerLys 956
 ::
 14537 AAAAGCGTAGTAAAGATTTTGAGTCTTCTTCAATGAAAAATCTCGAAA 14586
 957 ValGluGlyThrGluIleValLysProSerProLysArgLysMetGlu 973
 ::
 14587 CTAGAAGTCACTGAATATAGGAAGAACATCCAAAGCGCAAAATGGAAC 14636
 973 yAspValGluLysLeuGluArgThrProGluLysAspLysIleAlaSer 990
 ::
 14637 TGATACTGAAAAATGATAGAGACCCCTGAAAGCAAAAT...TCTT 14683
 990 eThrThrProAlaLysLysIleLysLeuAsnArgGluThrGlyLysLys 1006
 ::
 14684 TAGTCCGCCAGCCAAAAAATCAAACTCAGAGAGAAACTGGAGAGAA 14733
 1007 IleGlyAsnAlaGluAsnAlaSerThrThrLysGluProSerGluLys 1023
 ::
 14734 ATTGGAAGTACAGAAATATATATCAAAACACAAAGAACCCCTGAAAAAT 14783
 1023 uGluSerThrSerSerLysLysGluGluLysValLysGlyLysAla 1040
 ::
 14784 GGAGTCAACCTTAGCAAAAGTTAAACAGAAAAAGTCAAAAGAGTCA 14833
 1040 yAsArgLysValAlaGlySerGluGlySerSerSerThrLeuValAspTyr 1056
 ::
 14834 GAGCAAAAGTACGTGGAACCTGAAGAGATCCAGCTCACTGTGTGAGATTAC 14883
 1057 Thr.Ser..... 1058
 ::
 14884 ACCAGTAGCTGAGTAGAGGGGTGGGTGGAACCTTTGTGGAGAGAG 14933
 1058 1058
 14934 ATTTTCTTCAGTTTATTCATGCTGTGTCTTTTATATTAATATGTAAG 14983
 1058 1058
 14984 CTGATCTAGGAAGTAGCTGCTGTTTATATGAGCTTTTGTGTCTTGAA 15033
 1058 1058
 15034 GAAGGAAATGCCAAAGACAAGGTGTAAATAGATTGGGGGTGAGCCA 15083
 1058 1058
 15084 CATTTGTGTTGATACATCAATTAATCTTAATGTGTGTCTTCTTACGGC 15133
 1058 1058


```

XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
PA
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI: 2001-656860/75.
XX P-PSDB: ABB61740.
DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PR interactions -
PS
XX Claim 1: SEQ ID NO 12011; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
CC sequences (AB101840-ABL16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3910 BP; 1254 A; 961 C; 978 G; 717 T; 0 other;

alignment_scores:
Quality: 544.00 Length: 1405
Ratio: 0.890 Gaps: 55
Percent Similarity: 43.488 Percent Identity: 21.352

alignment_block:
US-09-811-045a-1 x ABL05843 ..

Align seg 1/1 to: ABL05843 from: 1 to: 3910

35 LysGluUysProPhleuProGluLu.....Pr 45
652 CCGGACAGCCAGATGCGCTGAGAACGATCAAGCATTTTGTGCTGCC 701
45 oSerSerSerSerGluGluAspAspProIleProAlaGluLeuLeuCysL 62
702 TGCTGTACAAACCAAGAG.....ATACCGAGAGATCTGATATGCG 742
62 eulleCysAspIleMetThrAspAlaValIleProCysCysGly 78
743 GCATATGCCAGATATATTCGTGATGCTCATGATACCTCTGCGGA 792
79 AsnSerSerCysAspGluCysIleArgThrThrLeuLeuGlnSerAspLy 95
793 AGTTCTTTTGTACGACTGTGCGAACCTCTTATTTGAGTCAAGAGA 842
95 shiThrCysProThrCysHisGlnAsnAspValSerProAspAlaLeuI 112
843 TAGTGAATGCCCGCATGCAAGAGAACTGTTGCTGCGCTCCCTGA 892
112 leAlaSnLysPheLeuArgGlnAlaValAlaSnAsnPhelysAsnGluThr 128
893 TACCTAATCGGTTCTTGAGGAATTCGGTGAACGCTTTAAATAATGAGACT 942
129 GlyTyrThrLysArgLeuArgLysGlnLeuProPhleuPheLeuVa 145

```

```

943 GGGTATATACAAAGCGGCTAG..... 966
145 lProProProArgProLeuSerGlnArgAsnLeuGlnProArgSerArgS 162
967 .....CCAGCTGCACTAAATAAT.....GAGGAAAGAC 994
162 eProIleLeuArg.....GlnGlnAspProValAlaPheArgTyrThr 176
995 CTCCTGTGAAAAAGAGAGTGAAGAAAAAGCCAGTCGG.....GAG 1035
177 ValSerProThrCysSerAspThrLysThrAlaGlySerCysSerAspSe 193
1036 GTGAACCCGAGAGAGACTGAGTGAAA..... 1062
193 rGlyThrLeuSerArgLeuProAlaProSerIleSerSerLeuThrSera 210
1063 .....CCTGAAAAAGAAAAAGATCCGAAACCAAT. 1092
210 snGlnSerSerLeuAlaProProValSerGlyAsnProSerSerAlaPro 226
1093 .....GGCAGTAATCCGCCAAATCCGAA 1116
227 AlaProValProAspIleThrAlaThrValSerIleSerValHisSerI 243
1117 TCTCCAGAGCTCCCGCAGACACAGAACATCA...CAGAAGAGAAAGA 1163
243 uLysSerAspGlyProPheArgAspSerAspAsnLysLeuLeuProAla 260
1164 TAAATATGATTCACTACACAGATTAACATTACCATTAATAATGCCAGC 1213
260 lAlaLeuThrSerGlnHisSerLysGlyAlaSerIleAlaIleThr 276
1214 CTGAGCTGATTCT..... 1227
277 AlaLeuMetGluGluLysGlyValProGlyThrSerProTrpAsnSerI 293
1227 ..... 1227
293 ePheValGlyGlnSerLeuLeuHisGlyGlnLeuIleProThrThrGly 310
1228 .....ACAAACATGC 1237
310 roValArgIleAsnAlaAlaArgProGlyGlyArgProGlyTyrGlu 326
1238 CCAGCAAAAGATCC.....CCAGTTATTCC 1263
327 HisSerAsnLysLeuGlyTyrLeuValSerProProGlnGlnIleArg 343
1264 CACAGAAAGTAAATCTCTCAT.....CG 1286
343 gGlyGluArgSerCysTyrArgSerIleAsnArgGlyArgHisHisSerG 360
1287 ACGGAGAGCTGATATGTTTCCGATCAGATCAACAAACAC.....C 1330
360 luarSerGlnArgThrGlnSerProSerLeuProAlaThrProCysPhe 376
1331 AACCTCATTAATAATCGAGCTGTTTACAAAGATCCAGT..... 1371
377 ValProValProProProLeuTyrProProProProHisThrIleUpr 393
1372 .....CTCCGACC 1379
393 oLeuProGlyValProProProGlnPheSerProGlnPheProSers 410
1380 CTTCGCCATTGGCACCTCGCTAGCTACAGGCGCACATGATGCGCAAT 1429
410 eGlnProProThrAlaGlyTyrSerValProProPro..... 422
1430 CAGAAGAGCTGTCGATCGAGTGCCATTAAAGCCCTTTATGCAAAATG 1479
423 .....GlyPheProProAlaProAlaAsnIleSerThrAlaCysPhe 437

```

1480 CAGCGGGGGCCACCCTCT.....ATGCACATGATGAG 1511
437 rProGluValProThrAlaHisSerAsnThrMetProThrThrGlnAlaP 454
1512 TCACCCACATCCAGCCCTACACAAAGGGTTTAAACACATGGGACAGAGGC 1561
454 roLeuLeuSerArgGluGluPheTyrArgGluGlnAsnAspLysGlyArg 470
1562 CTCGCCCTCAGC.....TATGTGCCGTATCAAAACCAATCCGTACAC 1602
471 GluSerLysPheProTyr.....SerGlySerSerTyrSerArgSe 484
1603 CCAATGGCGTCCGCCGTACGAGTCTGCAGGCGGAGGTATGAATATGATAT 1652
484 rSerTyrThrAspSerSerGlnGluLeuAlaGlnHisIleHisAlaLeuT 501
1653 GTCCACAAACCAATTTCAGTCCCAATTTAGCCTCGATATACCAAGGGGTGG 1702
501 hrLeuSerProSerAlaAlaHisThrLeuAspLeuLeuHisAspHisPro 517
1703 CAGCGAAGGTGGGTCCGGTCCGATTCACATCCGTTGGAGGCCCTTCAT 1752
518 HisProProGluGlnAlaGluAlaArgSerAlaMetIleValHisMetPr 534
1753 CGCATCTATGAAGGAGAGAGAGCGGAGAGAGTGCACCGCTTTCGAACTC 1802
534 oAspLeuMetAspIleAlaHisAlaArgSerArgSerProProTyrArgA 551
1803 TGAC.....CGCCACAGGTCAAGTCCCGGATGAGACAA 1837
551 rGlyTyrArgSerArgSerArgSerProProGluPheArgGlyGlnSerPro 567
1838 GGCACCCCTTTAAGTCT..... 1854
568 ThrLysArgAsnValProArgGluGluLysGluArgGluTyrPheAsnAr 584
1855CCCATGTACGAAAGACACATCC..... 1878
584 gTyrArgGluValProProProTyrAspIleLysAlaIleTyrTyrGlyArgS 601
1878 1878
601 erValAspPheArgAspProPheGluLysGluArgTyrArgGluTyrGlu 617
1879AGGGATTAATCTCAAGCACAACAAACCCGATCCCGGAA 1917
618 ArgLysTyrArgGluTyrTyrGluLysTyrTyrLysGlyTyrAlaValAl 634
1918 AGGAAGCGAGACATAGCTACGAAAGCATATACG..... 1953
634 yAlaGlnProArgProSerAlaAsnArgGluAspPheSerProGluArgL 651
1954CACCTCGTCTAGTCCGACGCCAATGATGGCTCT..... 1989
651 euLeuProLeuAsnIleArgAsnSerProPheThrArgGlyArgArgGlu 667
1990AAGTCCCAAGGTGGCAGATCAAAAGA... 2016
668 AspTyrAlaAlaGluGlnSerHisArgAsnArgAsnLeuGlyGlyAsnTy 684
2017TCTGGACATGCT...CGCTCTGCATCTCCAAAGCC 2048
684 rProGluLysLeuSerThrArgAspSer...HisAsnAlaLysAspAsn 700
2049 GGGCTACAAAGAGTGTATACAGAGACAAACCGTACAAAGCCTATAGCTC 2098
700 rOlySerLysGluLysGluSerGluAsnValProGly..... 712
2099 CCAAAAGCGAGCAGTGTAGCCCTCT...CCCCCGGATTCGAGCCGTTG 2145
713AspGlyLysGlyAsnLysHisLysLysHisArgGly 724
2146 CAGCTGACGAGTGAAGACGGCTACAGAAACAGACCCGACAGTTCCGGA 2195

724 sArgArgAsnGluGluLysGlyGluGluSerGluSerPheLeuAsnProG 741
2196 AGCATCAACAAGCAGCAGAGGGTGTATGACAC..... 2226
741 luLeuLeuGluThrSerArgLysCysArgGlySerSerGlyIleAspGlu 757
2227AAGAGAGAGGGGCAAAACAGCAGCAAGAG 2256
758 ThrLysThrAspThrLeuPheValLeuProSerArgAspAlaThrPr 774
2257 GCGCACGA.....AAGAGCGACAGTCTCGCAG 2285
774 oValArgAspGluProMetAspAlaGluSerIleThrPheLysSerValS 791
2286 CATTAGCAGAAACCGAAGCCGAATGACAC...AAGTACAGAGCCTGA 2332
791 erAspLysAspArgGluLysAspLysProLysValLysSerAspLys 807
2333 CTCACACAGCAAG...ATCACACACCGAAATGACTGCTGCCAG 2376
808 ThrLysArgLysSerAspGlySerAlaThrAlaLysLysAspAsnVal 824
2377 TTGAGCGCA..... 2385
824 uLysProSerLysGlyProGlnGluLysValAspGlyAspArgGluLysS 841
2386CGCCAAAGTT 2395
841 erProArgSerGluProProLeuLysLysAlaLysGluGluAlaThr... 856
2396 CACCGAAGAGC...CCGGAAGAGTCCAGCAGATATTCATCCCGCG 2439
857LysIleAspSerValLysProSerSerSerSerGlnLysAspG 871
2440 AAGCCAGAAATTATGAGCTCCACGCCGTCAATCAACACAGCAAGTAA 2489
871 uLysValThrGlyThrProArgLysAlaHisSerLysSerAlaLysAsp 887
2490 GACCAATGTGGCAGAGAGACAAAGGCCAGAGTCCGTTGCAAAAGATC 2539
888 ThrArgArgGlnSerGlnProArgThrArgArgSerLysArgThrValPr 904
2540 GCAGAAAGAAAGAGAGAGACAAAGCAGCTGAGCCCAAG... 2581
904 oLysThrSerSerGlnLysSerGlnProValArgThrArgArgProArgS 921
2582AAAACAGAAAGAGACAAAGC 2600
921 erLeuArgLysIleAsnTyrLeuIleAlaArgGluLysAsnGluArgGlu 937
2601 CGCTAAGAG.....GAGAAAGGGATCGCCAG 2628
938 LysArgLysLysSerVal.....AspLysAspPheGluSerSerSe 951
2629 AAGAAGAGCTCTCAGTTAATCGATCTGACTCGGATATTACACAGCTC 2678
951 rMetLysIleSerLysValGluGlyThrGluIleValLysProSerProL 968
2679 ACTAATG..... 2685
968 ySArgLysMetGluGlyAspValGluLysLeuGluArgThrProGluLys 984
2685 2685
985 AspLysIleAlaSerSerThrThrProAlaLysLysIleLysLeuAsnAr 1001
2686 2685
1001 gGluThrGlyLysLysIleGlyAsnAlaGluAsnAlaSerThrThrLysG 1018
2686AACGAGTCAATATATTAAG 2704

```

1018 1upSerGluLysLeuGluSerThrSerSerLysIleLysGlnGluLys 1034
      ||| : : : ||| ||| : : : ||| : : :
2705 TATTGCTCCCAAGGGCTCAAGTCCAGCATGATGATCAATGGGCTCA 2754
      ||| : : : ||| : : : ||| : : : ||| : : :
1035 ValLysGlyLysAlaLysArgLysValAlaGlySerGlySerSerse 1051
      : : :
2755 CTT..... 2757
1051 rThrLeuValAspTyrThrSerThrSerSerThrGlySerProVala 1068
      ||||| : : :
2758 .....TCCCTTATC 2767
1068 rGlySerGluGluLysThrAspThrLysArgThrValIleLysThrmet 1084
      : : : : : ||| : : : ||| : : :
2768 ACAACCGCTACTGAAACGTTAATCCGAAAGATCCATTCCTTACTGTG 2817
      : : : : : ||| : : : ||| : : :
1085 GluGluTyrAsnAsnAspAsnThrAlaProAlaGluAspValIleIle 1101
      : : : : : ||| : : : ||| : : :
2818 GGTGCTGACGAGCATATCTTGCCCAAGAACAACTCAGCGAGGC 2867
      : : : : : ||| : : : ||| : : :
1101 tIleGlnValProGlnSerLysTrpAspLysAspAspPheGluSerGlu 1118
      : : : ||| ||| ||| ||| : : :
2868 TAATTCTGTCAATCTATCCAAATGGGAAATCGACGAAATATCTTAGCTT 2917
      ||||| : : : : : ||| : : : ||| : : :
1118 1uGluAspValLysThrThrGlnProIleGlnSerValGlyLysProser 1134
      : : : : : ||| : : : ||| : : :
2918 TGGAGGAT...TCCTCCAAAAAGCTGCGGCGCTCCGACGATCCGCTG 2964
      : : : : : ||| : : : ||| : : :
1135 SerIleIleLysAsnValThrThrLysProSerAlaThrAlaLysTyrTh 1151
      : : : ||| ||| : : : ||| : : :
2965 GAAATATCACTCAAGCTCCTCGCAAAAGGCTGAGAACGCAATATTTCGAAA 3014
      : : : : : ||| : : : ||| : : :
1151 rGluLysGluSerGluGlnProGluLysLeuGlnLysLeuProLysGlu 1168
      : : : : : ||| : : : ||| : : :
3015 GCGTATTATGCGCATCAGCGCTATGAGATTCAAGTTATT..... 3054
      : : : : : ||| : : : ||| : : :
1168 1aSerHisGluLeuMetGlnHisGluLeuArgSerSerLysSerAla 1184
      : : : : : ||| : : : ||| : : :
3054 ..... 3054
1185 SerSerGluLysGlyAlaLysAspArgGlnHisSerLysGluLys 1201
      : : : : : ||| : : : ||| : : :
3055 ATCAATTCACAGGACAAAGCAAGGACCGCTCCGATTCGAAAGTGACAA 3104
      : : : : : ||| : : : ||| : : :
1201 sAsp.....AsnProAspLysArgLysSerGlyAlaGlnProAspL 1215
      |||| : : : : : ||| : : : ||| : : :
3105 GGATCGCTCTCTCTCACCCAGGCGGTAAACAACG.....A 3139
      : : : : : ||| : : : ||| : : :
1215 ysgLysSerThrValAspArgLeu.....serGluGlnGly 1226
      : : : : : ||| : : : ||| : : :
3140 GCAGGTCGGTAAAGATAGGCTGGGCACCAAGATTCCCAATGATAGAAC 3189
      : : : : : ||| : : : ||| : : :
1227 HisPheLysThrLeuSerGlnSerSerLysGluThrArgThrSerGlu 1243
      : : : : : ||| : : : ||| : : :
3190 CGTTCGCGAGACAAGTCGAAGGCGAGCGCGCGCAAGAGCTCCGA 3239
      : : : : : ||| : : : ||| : : :
1243 sHisGluSerValArgGlySerSerLysAspPheThrProGlyAlaArg 1260
      : : : : : ||| : : : ||| : : :
3240 CGACGATGGCAACCGCGAGCTCGATGCGACGCGGAGAGAGG 3289
      : : : : : ||| : : : ||| : : :
1260 sPlysLysValAspTyrAspSerArgAspTyrSerSer...SerLysArg 1275
      : : : : : ||| : : : ||| : : :
3290 ACAACAGA.....TCCCGGACACAGGCGCGCTTACAGAAC 3327
      : : : : : ||| : : : ||| : : :
1276 ArgAspGluArgGlyGluLeuAlaArgAlaLysAspSerProAlaArg 1292
      : : : : : ||| : : : ||| : : :
3328 AGCCAGAGAGCGT.....TCGTACAAGCGCAAGCTCCGCG..... 3360
      : : : : : ||| : : : ||| : : :
1292 yLysGluSerLeuSerGlyLysLysSerLysLeuArgGluGluArgAspL 1309
      : : : : : ||| : : : ||| : : :
3361 .....GAGGACGACAAAGCTGAGGCGCGCAAGAAC.... 3387
      : : : : : ||| : : : ||| : : :
1309 euProLysGlyAlaGluSerLys.....LysSerAsnSer 1321

```

```

3388 .....AAGAACAGTCCGAATCCAAACGACGAAAGCATGATCAAAACAT 3432
      ||| : : : ||| ||| : : : ||| : : :
1322 SerProProArgAspLysLysProHisAspHisLysAlaProTyrGluTh 1338
      ||| : : : ||| : : : ||| : : : ||| : : :
3433 AGCAGACACTCGGATCGAGGGCGGCGCAAAACACCAAGTCCAGCGACAG 3482
      : : : : : ||| : : : ||| : : : ||| : : :
1338 rLys.....ArgProCysG 1343
      : : : : : ||| : : : ||| : : : ||| : : :
3483 CCGAGTGGTCTCTCTGTAAACAGCGCTGCTCTCCCAACCTGTG 3532
      : : : : : ||| : : : ||| : : : ||| : : :
1343 1uGluThrLysPro.....ValAspLysAsnSerGlyLysGlu 1355
      : : : ||| ||| : : : ||| : : : ||| : : :
3533 GTCCACACACACACCGCTCCGCAAGTTGTCGACACCGAGTTGTCGACACAG 3582
      : : : : : ||| : : : ||| : : : ||| : : :
1356 Arg.....GluLysHisAlaAlaGluAlaArg 1364
      : : : : : ||| : : : ||| : : : ||| : : :
3583 TTAGTTGTAAATATGATATAACAGATACAGAAAGAGGCGCGTCTCGGA 3632
      : : : : : ||| : : : ||| : : : ||| : : :
1364 gAsnGlyLysGlu 1368
      : : : : : ||| : : : ||| : : : ||| : : :
3633 CAACGGCATGAG 3645
      : : : : : ||| : : : ||| : : : ||| : : :
seq_name: /SID51/gcdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABL05842
seq_documentation_block:
ID ABL05842 standard; cDNA: 7385 BP.
XX
XX ABL05842;
XX
XX 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 12008.
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB61739.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 12008; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX sequences (AB101840-AB16175) and the encoded proteins
XX (AB57737-AB572072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WPI
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX

```

SQ Sequence 7385 BP; 2213 A; 1708 C; 1645 G; 1819 T; 0 other;

```
alignment_scores:      530.50      Length:      1513
                       Quality:
Ratio:      0.792      Gaps:      67
Percent Similarity:    44.283      Percent Identity:    20.952
```

```
alignment_block:
```

Align seg 1/1 to: ABL05842 from: 1 to: 7385

35 lysGluIysProProPheLeuProGluGlu.....Pr 45
 2594 CGGACAAAGCCAAATATCGGCTGGAAGCAATACGCCGATTTGTGCTGC 2643
 45 oSerSerSerSerGluGluAspAspProIleProAlaGluLeuLeuCyl 62
 2644 TGCTGTACAAACCAAGAG.....ATACGGAGAGATCTATATGCC 2684
 62 euIleCysIysAspIleMetThrAspAlaValIleProCysCysGly 78
 2685 GCATATGCCAGATATATATGCTGGATGCTGTATGATACCTCTGGCGGA 2734
 79 AsnSerSerCysAspGluCysIleLeuGlyThrLeuLeuGluSerAsp 95
 2735 AGTTCCCTTTGTGACACGCTGTGTCGCAACCTCCTTATGGAGTCAGAGA 2784
 95 sHsIThrCysProThrCysHisGlnAsnAspValSerProAspAlaLeu 112
 2785 TAGTGAAGTCCCGGATTGCAAGGAAAGAACTGTTGCTTGCTCCTTGA 2834
 112 leaIAsnIysPheLeuArgGlnAlaValAsnAspNheIysAsnGluThr 128
 2835 TACCTATTCGGTTCCTTGAGGAATTCGGTGAACGGCTTAAATAATGACAT 2884
 129 GlyThrThrIysArgLeuArgIysGlnLeuPro.....139
 2885 GGGTATATACAAAGCGCGGCTAAG.....CCAGGTAGTTCAATATATCT 2928
 140ProPheLeuPheLeuValPro 147
 2929 CGTTTGCATCGCTTAATGCAAAATGATCGAATTTCTCTTTTACTGCA. 2977
 147 roProArgProLeuSerGlnArgAsnLeuGlnProArgSerArgSerPro 163
 2978GTAATAATGAGCAAAAAACCTCCT 3001
 164 IleLeuArg.....GlnGlnAspProValValPheArgThrValSe 178
 3002 GTTGAAAAAGAAAGTGGAGAAATAACCAAGTCGCG.....GAGTGGGA 3042
 178 rProThrCysSerAspThrIysThrIleArgIysSerCysSerAspSerGly 195
 3043 ACCCGAAGACACTGAGGTGAAA..... 3064
 195 hrlLeuSerArgLeuProAlaProSerIleSerSerLeuThrSerAsnGln 211
 3065CTGTAAAGCAAAAAGATCCGAATCCAAAT..... 3094
 212 SerSerLeuAlaProProValSerGlyAsnProSerSerAlaProAlaPr 228
 3095GGCAGTATATCCGCAAAATGCGAATCTCC 3123
 228 oValProAspIleThrAlaThrValSerIleSerValHisSerGluIys 245
 3124 AGAGCTCTCCCAACACAGAACCATCA...CAGAGAGAGAAAGATTAAT 3170
 245 eraSpGlyProPheArgAspSerAspAsnIysLeu.....ProIa 259
 3171 ATGATTCAGACTACGAGATTAACATTACCAATAAAATCCCAAGCTCTCA 3220

[illegible]

```

3797 AAGAGTAATTAAAGCTGGGCTGCGCCTTATGCTTTGTTGAATATT      |||::
      ::::::::::      |||::
531 LHisMetProaspLeuMetAspIleAlaHisIaIarSer..... 544
      ::::::::::      |||::
3847 GGTATTTCTTAATTAAACAGATTGCTTACATCAACGTTAAAGCTTTT 3896
545 .....ArgSerProProTyrArgArgTyrArgSerArg 555
      ::::::::::      |||::
3897 GAATTGTATTACTTCTAAATTTACAGTATGTCGCCGATCAAAACCA 3946
556 SerArgSerProProGluPheArgGlyGlnSerProThrLysArgAsnVa 572
      |||
3947 TCCGTACACCA..... 3958
572 LProArgGluGluLysGluArgGluTyrPheAsnArgTyrArgGluValP 589
      ::
3959 .....ATGC 3962
589 roProProTyrAspIleLysAlaTyrTyrGlyArgSerValasPheArg 605
      |||||
3963 GTGCCCGCGTAACGA...TCTGCAGCGCGAGTATGATATGATATGATCA 4009
606 AspProPheGluLysGluArg.....Tyr.ArgGluTyrGluVal 618
      ::::::::::      |||
4010 CAACCATTTCACTCCCAAAATTTAGCTCGATATACCAAGGCGTGACAC 4059
618 rGlys.....TyrArgGluTyrPyr 624
      |||
4060 GAAGGCTCGCTCCGGGTGAGTATCTTTAGGCGCTTGAGAACGCGTGCGAT 4109
625 GluLysTyrTyrLysGlyTyrAlaValGlyAlaGlnProArgProSerAl 641
      ::::::::::      |||
4110 GACAAAGTATCCCAATGTTTGCAGTC..... 4136
641 aAsnArgGluAspPheSerProGluArgLeuLeuProLeuAsnIleArg 658
      |||||
4137 .....CCATTGACGATCCGTT 4152
658 snSerProPhe.....ThrArgGlyValArgArg...GluVal 668
      |||
4153 GAGGCGCTTCAATCCCATCATGACAGAGAACGAGGAGAGGTGGAC 4202
668 sPtyrAlaAlaGlyGlnSerHisArgAsnArgAsnLeuGlyLysAsnTyr 684
      ::::::::::      |||
4203 GCTTCGAAAGCTGACCGCCACAGGTCAAGTCC..... 4237
685 ProGluLysLeuSerThrArg.....AspSerHisAs 695
      |||
4238 CCGGATAGCAAAAGCACCGCTTTAAGTCTCCCATGTACGAAAGACAA 4287
695 nAlaLysAspAsnProLysSerLysGluLysGluSerGluAsnValProG 712
      ::::::::::      |||
4288 CTCGAGGGATATCTCAAGAGACAAAGACCGCGATCCCGGAAAGAGAAC 4337
712 LysAspGlyLysGlyLysAsnLysHisLysLysHis...ArgLysArgArgAsn 727
      ::::::::::      |||
4338 GAGAACATAGCTAAGACGCGCATATACGCCACCTCTCTAGTGGCCAG 4387
728 GluGluLysGlyGluGluSerGluSer..... 736
      ::::::::::      |||
4388 CCGAATGATGCTTAAGTCCCAAGGTGCGAAGATCAAAAGTAGTTTAA 4437
737 .....PheL 738
      |||
4438 CCACCAAGCCCTTTCCATGTGCTTACATTCATGCTCTCTCTCTTTT 4487
738 euAsnProGluLeuLeuGluThrSerArgLysCysArgGlySerSerGly 754
      ::::::::::      |||
4488 TAGATCTGGACATCGTGGCTGTGATCTCCAAAGCGGCTACAGAGTGC 4537
755 IleAspGluThrLysThrAspThrLeuPheValIleuProSerArgAspAs 771
      |||

```

```

4538 ATTACAGACAGAACCGCTACAAAGCCTAGTGTCCCAAAACGAG... 4584
771 palatThrProValArgAspGluProMetAspAlaGluSerIleThrPheL 788
      |||
4585 .....GCAGTTGAGCCTCTCTCCCGGATTCGAGCCGTGG..... 4620
788 ySerValSerAspLysAspLysArgGluLysAspLysProLysValLys 804
      ::::::::::      |||
4621 ..CAGCTGACGATGACAGCGGCTACAGSACACACCGCACAGGTTGC 4668
805 SerAspLysThrLysArgLysSerAspGlySerAlaThrAlaLysLysAs 821
      ::::::::::      |||
4669 GAAGCATCAAAAGCAGCAGCGGTGTAGCAGC.....AAGAAGAG 4709
821 pasnValLeuLysProSerLysGlyProGlnGluLysValAspGlyAspA 838
      ::
4710 AGGGAAAAACAGCACAGAGAGCGCCACGAAAGAG.....CACA 4750
838 rGluLysSerProArgSerGluProProLeuLysLysAlaLysGluGlu 854
      |||
4751 GGTCTCGGAGCATTAGCAAGAACCG..... 4776
855 AlaThrLysIleAspSerValLysProSerSerSerSerGluLysAsp.. 870
4777 .....AAGCGAATGACACCACTACAGAGAGCT 4805
871 .....GluLysValThrGlyThrProGlyLysAlaHisSerLys 884
      |||
4806 GACTCCACCAGCAAAAGATCAC...ACACCAAAATGACT.....G 4843
884 eAlaLysAspThrArgArgGlnSerGlnProArgThrArgArgSerLys 900
      ::::::::::      |||
4844 CTGCCAGATTGAGCAACGCAAGGTTTCACGACAGCGCGGAAAAAGAT 4893
901 Arg.....ThrValProLysThrSerSerGlnLysSerGlnProVal 914
      ::::::::::      |||
4894 CACGAGATTATCTGACCGCAAGGCGCAAGTATATGCTCCACCCCT 4943
914 L.....ArgThrArgArgProArgSerLeuArgLysIleAsnTyrLeuI 929
      |||
4944 CATCAACGACACGGAATGAGAACCAATGTGGCAAGGAGAACAAAG... 4989
929 lAlaArgGluLysAsnGluArgGluLysArgLysLysSerValAspLys 945
      ::::::::::      |||
4990 ..GCCAAGAGTCCGTTGTCAAAAGATCGCAAGAGAAAGAGACAG 5037
946 AspPheGluSerSerSerMetLysIleSerLysValGluGlyThrGluI 962
      |||
5038 GAC.....AAGCGTACGCGCAAGAAAAA 5060
962 eValLysProSerProLysArgLysMetGluGlyAspValGluLysLeuG 979
      |||
5061 CAAGAAAGACAAAGCGCGCTAAGAGAGAAAGGATCCCAAGAG... 5106
979 LuArgThrProGluLysAspLysIleAlaSerSerThrThrProAlaLys 995
      |||
5107 .....AAG 5109
996 LysIleLysLeuAsnArgGluThrGlyLysLysIleGlyAsnAlaGlu.. 1011
      ::::::::::      |||
5110 ACCTCTCTCAATTATCGA...TCTGACTCGGATATTAAACAGCTCACT 5156
1012 ...AsnAlaSerThrThrLysGluProSerGluLysLeuGluSerThs 1027
      |||
5157 AATGAACGAGTCAAAATTAAAGTATGTCTCCAGGCGTCAAGTCCCA 5206
1027 erSerLysIleLysGlnGluLysLysGlyLysAlaLysArgLysVal 1043
      ::::::::::      |||
5207 GCATTGAGATCAATCGGCTCAACTT..... 5232
1044 AlaGlySerGluGlySerSerSerThrLeuValAspTyrThrSerThse 1060
      ::::::::::      |||
5232 ..... 5232

```


Percent Similarity: 88.194 Percent Identity: 71.528

alignment_block:
US-09-811-045a-1 x AAI82535 ..

Align seg 1/1 to: AAI82535 from: 1 to: 439

```

1041 ARGVSLVALAGLSERGLUGLSERSESRTHLEUVALASPTYRTH 1057
      |||||||.....|||||||||||||||||||||.....|||||
7 CGAAGAGTCTGAGACGAGATCTCACTCAAGTGTCTGGATTAC 56
1057 rserThrSerSerThrGlySerProValArgLysSerGluGluLysT 1074
      | |||||||.....| |||||||.....| |||||||
57 CATTCCGACTCACTGAGGAGCATGCTGTGCGGAATCTGAAGACCAA 106
1074 hAspThrLysArgThrValLLeLysThMetGluGluTyrAsnAsp 1090
      |||||||.....| |||||||.....| |||||||
107 CAGATTACTAAGCGAAGCTGATTAAACATGAGACACTATAATATGAC 156
1091 AsnThrAlaProAlaGluAspValLLeMetLLeGluValProGlnSe 1107
      |||||||.....| |||||||.....| |||||||
157 AATACCGCTCCAGCTGAGATGTTATCATTTATGATTCAGGTTCTCAATC 206
1107 rLysTPAspLysAspAspPheGluSerGluGluGluAspValLysThrT 1124
      | |||||||.....| |||||||.....| |||||||
207 CATATGGAATCAAGATGACTTTCATCTGAGACCAAGATGATAATCCA 256
1124 hGlnProLLeGlnSerValGlyLysProSerSerLLeLysAsnVal 1140
      |||||||.....| |||||||.....| |||||||
257 CACAGCCTATATCAAGTGTAGGAAACCTGCTCTATATAAAATGTT 306
1141 ThrThLysProSerAlaThrAlaLysTyrThrGluLysGluSerGluG 1157
      |||||||.....| |||||||.....| |||||||
307 CGTACAAAGCCATCTCTATAGTCAAGTATCCTGAGAAAGAAAGTAGGCC 356
1157 nProGluLysLeuGluLysLeuProLysGluAlaSerHISGluLeuMetG 1174
      |||||||.....| |||||||.....| |||||||
357 ATTCCGCTAAATTCGAAATTCACCATGAGCGTGAATGATGATGATC 406
1174 LhISGluLeuArgSerSerLysGlySerAla 1184
      |||||||.....| |||||||.....| |||||||
407 CTCATGAGGTTAAAGTTCAATATCTCTGCA 438
seq_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV21511
seq_documentation_block:
ID AAV21511 standard; DNA; 6755 BP.
XX
AC AAV21511;
XX
DT 17-AUG-1998 (first entry)
XX
DE Staphylococcal bacteriocin BacRI operon.
XX
KW BacRI; bacteriocin; antimicrobial; antibacterial; antibiotic;
KW Moraxella bovis; infectious bovine keratoconjunctivitis; cancer;
KW therapy; ds.
XX
OS Staphylococcus aureus strain UT0007 (ATCC 55800).
XX
PN MO9812319-A1.
XX
PD 26-MAR-1998.
XX
PE 18-SEP-1997; 97WO-US16758.
XX
PR 17-SEP-1997; 97US-0931999.
XX
PR 19-SEP-1996; 96US-0710561.
XX
PA (UNIV ) UNIV KANSAS STATE RES FOUND.
XX
PI Crupper SS, Iandolo JJ,
XX

```

DR WPI: 1998-230316/20.

XX Therapeutic proteinaceous substances from *Staphylococcus aureus* -
PT useful to inhibit growth of wide range of prokaryotic or eukaryotic
PT cells, e.g. *Moraxella bovis* causing infectious bovine
PT keratoconjunctivitis
XX
PS Claim 2; Page 19-23; 38pp; English.

CC This polynucleotide comprises the bacteriocin BacRI operon of
CC *Staphylococcus aureus* UT0007. The sequence of the BacRI operon
CC was determined by N-terminal sequencing of purified BacRI peptide
CC (see AAW54171), with back-translation and plasmid analysis. The
CC BacRI operon includes the BacRI gene (see AAV21510), a homologue of
CC the *cylM* gene of the cytolysin operon of *Enterococcus faecalis*
CC whose function is involved in the maturation of pre-cytolysin,
CC an ATP-transporter gene, *bio1* and *bio2* genes related to
CC lactococcal biosynthesis and modification, and a gene involved in
CC immunity function. BacRI peptides can be produced by construction
CC of an expression vector containing an oligonucleotide or operon
CC coding for BacRI, and use of the vector to transform host cells for
CC BacRI expression. The entire BacRI operon has been cloned into
CC plasmid pUB110, and *Bacillus subtilis* transformants secreted the
CC recombinant BacRI peptide into the medium. Bacteriocin BacRI is
CC active against many Gram-positive and Gram-negative organisms such
CC as *Bordetella bronchiseptica*, *Pasteurella multocida* and
CC *Staphylococcus aureus*; *Moraxella bovis*, causing infectious bovine
CC keratoconjunctivitis, is especially sensitive. BacRI can also be
CC used as an anti-cancer agent.
SQ Sequence 6755 BP; 3903 A; 1252 C; 1600 G; 0 U; 0 other;

alignment_scores:
Quality: 477.00 Length: 995
Ratio: 0.958 Gaps: 40
Percent Similarity: 50.050 Percent Identity: 21.508

alignment_block:
US-09-811-045a-1 x AAV21511 ..

Align seg 1/1 to: AAV21511 from: 1 to: 6755

```

550 ARGArgTyrArgSerArgSerArgSerProProGluPheArgGlyGlnSe 566
      ||||| |||.....| |||.....|
2890 CGAGCGCAACGAGAAAGAAACAGA..... 2913
566 rProThrLysArgAsnValProArgGluGluLysGluArgGluTyrPhea 583
      :||| |||.....| |||.....|
2914 .....AGAGAGAAC...CCGAGGAGAACGAAAGCAGCAAGCAGCA 2953
583 snArgTyrArgGluValProProTyrAspLLeLysAlaTyrTyrGly 599
      :|||
2954 GGAGA..... 2958
600 ArgSerValAspPheArgAspProPheGluLysGluArgTyrArgGluTr 616
      |||||.....| |||.....|
2959 .....GAGAAAAAGAACAGACAGACAC 2981
616 pGluArgLysTyrArgGluTyrPtyrGluLysTyrTyrLysGlyTyrAlav 633
      :||| |||.....| |||.....|
2982 CAGAAAGCCAGCAAAAAAGAAAAA.....GAGGGACACACAG 3025
633 aLysAlaGlnProArgProSerAlaAsnArgGluAspPheSerPro... 648
      :||| |||.....| |||.....|
3026 CAGGGGACGCGGACAGAAACACAGAAACAGAACCCGAGG 3075
649 .....GluArgLeuLeuProLeuAs 655
3076 CAGCAAGGGGGAGAGAAAAACAAAAACAGAAAAAGACACAAAGACAC 3125
655 nLLeArgAsnSerProPheThrArgGlyArgArg..... 666
      :|||.....| |||.....|

```

```

3126 CACAGAAAGCAACCCAAAGACAGAGAAAGAGGGGCAACACAGAAA 3175
667 .....GluAspTyr 669
3176 CGAAAAACAAACGGCCACAGCCAAACAAAGCCGCAAGGAAAGACAGA 3225
670 AlaAlaGlyInSerHisArgAsnArgAsnLeuGly ..... 681
3226 GCGAAAGGGCAGCAGAGGACAAACGACAAAGAAACAAAGCAAAAGCCGC 3275
682 .....GlyAsnTyrPro GlyLysLeuSerThrArgAspSer 693
3276 AACCCCAAAAGAGAGAGCAACCCAGAAAGGCGCAGACACAGAGAGAGAA 3325
694 HisAsnAlaLys...AspAsnProLysSerLysGluLysGluSerGluAs 709
3326 CACGAGAAAGAAAGCCAAACGAAAGCCAGCAGCAAGAGAGAGAAAG 3375
709 nValProGlyAspGlyLysGlyLysHisLysLysHisArgLysArg 726
3376 GAGGACAGAGAAAGAAAGGCGCAGCAAGAAAGAAAGAAAGGAAAGAAA 3425
726 rgaAsnGluGluLys.....GlyGluGluSerGluSerPheLeuAsn 739
3426 ACAGAGCGAAAAAGAGAGGAGAGAGAGAAAGAAAGAGACAAAAAAC 3475
740 ProGluLeuLeuGluThrSerArgLysCysArgLysSerSerGlyLeas 756
3476 ACCGAGCGAGAAAGAAAGAAAGGAAAGAAAGAAAGAAAGAAAGAAACA 3525
756 pGluThrLysThrAspThrLeuPheValLeuProSerThrArgAspAla 773
3526 GAAAGACAAAGAGAGA..... 3541
773 hrProValArgAspGluProMetAspAlaGluSerIleThrPheLysSer 789
3542 ..GAAAAAAGAAAGAAACAGACAGCAAGCAAGAAACAGCGGAAACAC 3589
790 ValSerAspLysAspLysArgGluLysAspLysPro..... 801
3590 GAAAAAAGCGAGACAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGCGCAG 3639
802 .....LysValLysSerAspLysThrLysArgLys 812
3640 CGACCAAAAAAGAAAGCCGGAACCAAGCAAGAAAGAAAGCAAGAGAGA 3689
812 eraSpGlySerAlaThrAlaLysLysAspAsnValLeuLysProSerLys 828
3690 AAGAAAAAAGAAAGAAAGCGAGCAAAA.....GCCGACGAAAGAAC 3727
829 GlyProGluGluLysValAspGlyAspArgLysSerProArgSerGlu 845
3728 GGAGGAAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGCGCGGA 3777
845 uProProLeuLysLysAlaLysGluGluAlaThrLysIleAspSerVal 862
3778 CAGACCAAGAAAGAAAGGAGAGAGAGAAAGCAAGAAAGAAAGAAAGAAC 3827
862 ysProSerSerSer.....SerGluLysAspGluLysValThrGly 875
3828 AAGAGAGAGAGCAAGAAAGAAAGAAAGAGAGAGCAAGAAAGAAAGCCGC 3877
876 ThrProArg.....LysAlaHisSe 882
3878 ACAAGAGAGAGAGAGAGAGAGCAAGCAAGAAAGCAAGAAAGCAAGAGAC 3927
882 rLysSerAlaLysAspThrArgArgLys..... 891
3928 CAAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAC 3977
892 .....SerGluProArgThrArgArgSerLys 900
3978 AAAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4027

```

```

901 ArgThrValProLysThrSerSerGluLysSerGluProValArgThrArg 917
4028 GAACACGCGGAAAAAACCAAGCAGACACAAACCAAGAAAGAAAGAGACAG 4077
917 gArg.....ProArgSerLeuArgLysIleAsnTyrLeu 929
4078 AAGAGAGGCAAAAGCCCAACCCCAAGAAAGAAAGAAAG..... 4114
929 LeuAlaArgGluLysAsnGluArgGluLysArgLysLysSerValAspLys 945
4115 ..GCCGCAAAAGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4150
946 AspPheGluSerSerMetLysIleSerLysValGluGluThrGlu 962
4151 .....GACCGGAACCAAAAGCAAAAGCAAGAGAGAGAGAGAGAGAGAG 4191
962 eValLysProSerProLysArgLysMetGluGlyAspValGluLysLeu 979
4192 AACCCACCCACAA..CGACAGAAAGCAAAAGAG.....AAGCGGCGG 4232
979 LuArgThrProGluLysAspLysIleAlaSerSerThrThrProAlaLys 995
4233 AAAAGCAGACCAAAAGAAAGCAAA.....AACACACAAAGAGAGAGAAA 4273
996 LysIleLysLeuAsnArgGluThrGlyLysLysIleGlyAsnAlaGluAs 1012
4274 AAG.....GCCACGACAAAGCAAGCCGAGCAAGAAAGAAC 4305
1012 nAlaSerThrThrLysGluProSerGluLysLeuGluSerThrSerSer 1029
4306 CCAGAAAGCAAAAGAAAGAAAGAAAGCAAGAAAG.....AAGCAGCAAGAGCC 4349
1029 ysIleLysGluGluLysValLysLysLysAlaLysArgLysValAlaGly 1045
4350 AACCAAAAGAAAGAAAGAAAGCCGAGAAAGAAAGAAAGAAAGAAAGAAA 4399
1046 SerGluLysSerSerSerThrLeuValAspThrThrSerThrSerThr 1062
4400 CAAAGAGAAACCCCAACAAAC.....AGCAAAAGAAAGCAAGAGAGAGAC 4443
1062 rGlyLysSerProValArgLysSerGluGluLysThrAspThrLysArg 1079
4444 AGGAAAGCGAGAACCCAGAGAGAAAGAAAGGAGAAAGAAAGCAAAAGAAA 4493
1079 hrValIleLysThrMetGluGlu..TyrAspAsnAspAsnThrAlaProAl 1095
4494 GAAAGCAAAAGCAAAAGCGAGAGCAAGCAACACAAAGAGCGGAGAC 4543
1095 aGluAspValIleIleMetIleGluValProGluSerLysTrpAspLys 1112
4544 GGAAGAA.....AAACACCCCGAAGAAAGAAAGCAAGAAAGAG 4578
1112 spAspPheGlu..SerGluGluGluAspValLysThrThrGluProIleGlu 1128
4579 ACAACACAAAGCAAAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4620
1128 nSerValGlyLysProSerSerIleIleLysAsnValThrThrLysPro 1145
4621 .....AAAAAGAAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4642
1145 eraAlaThrAlaLysTyrThrGlu..... 1152
4643 AGGAAAGAGCCAGAGAAAGCAAGCAAGCAAGAGAGAGAGAGAGAGAGAG 4692
1153 LysGluSerGluGluProGluLysLeuGluLysLeuProLysGluAlaSe 1169
4693 AAAAGAGAGAGCAAAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4742
1169 rHisGluLeuMetGlnHis..... 1175
4743 ACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4792

```

```

1175 ..... 1175
4793 CAAAACGAAAGACGACAAAGGAGGCGCCGACGACCAAAAGAAAGAA 4842
1176 ..... GlutEArgSerSe 1180
4843 GAGAGACCCAGAGAGAGACACCCCGACACCCGACAAAGAAAGAAAG 4892
1180 rlysglyseralaserseglu...lysglyargalaalysaparglu 1196
4893 ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4941
1196 lsserglyserglulysaspasnproasplysarglylsergluln 1212
4942 ..GCAAAACGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4989
1213 proasplysglulserthrvalaspargluergluinllyshisphely 1229
4990 AAACACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5028
1229 sthrleuSerGluSerlyserlysgluThrArgThrSerGluLysHis 1246
5029 .....GACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5071
1246 ervalarglyserSerAsnLysaspserthrprogllyargasplylys 1262
5072 CCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5115
1263 valasplyrserasparasplyrserSerlyserlyarg.....Ar 1276
5116 CAGAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5165
1276 gaspblarglygluleualaarg.....Lysasps 1288
5166 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5215
1288 erproasplyrlysglyserleuSerGlyLysSerlyLysLeuArg 1304
5216 AAACGCGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5265
1305 glulguarglyleuProLysLysGlyAlaLysSerlyLysSerAsnSe 1321
5266 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5315
1321 rSer.....ProPThrArgAspLysLysProH 1330
5316 AACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5365
1330 lssasphs.....LysAlaProTyGluThr 1338
5366 AAACGAGACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5415
1339 lysarprocygslugluThrLysProValasplysasnserGlyLysG 1355
5416 AGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5465
1355 uargglulysHisAlaala.....G 1362
5466 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5515
1362 lualaargasnlyLysgluSerGly 1371
5516 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5544

```

seq_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA251264

seq_documentation_block:

ID AA251264 standard; cDNA; 680 BP.

XX AA251264;

XX 06-JUN-2000 (first entry)

XX

```

DE Human RNA-associated protein-15 (RNAAP-15) encoding cDNA.
XX
KW RNA-associated protein; RNAAP; human; clone 1879133; cytosolic;
KW immunosuppressive; antiinflammatory; keratolytic; neuroprotective;
KW antiarteriosclerotic; hepatotropic; antiposoriatic; viticude; anti-HIV;
KW antiallergic; antirheumatic; antiarthritic; ophthalmological; autoimmune;
KW antimicrobial; cell proliferative disorder; inflammation; cirrhosis;
KW actinic keratosis; bursitis; arteriosclerosis; artherosclerosis;
KW hepatitis; myelofibrosis; primary thrombocytopenia; psoriasis; cancer;
KW mixed connective tissue disease; MCTD; HIV; uveitis; Crohn's disease;
KW allergy; rheumatoid arthritis; parasitic infection; ss.
XX
OS Homo sapiens.
XX
FH key Location/Qualifiers
FT CDS 117..578
FT /tag= a
FT /product= "Human RNA-associated protein-15"
FT /note= "Derived from LEUKNOT03 library"
XX
PN MO20001171-A2.
XX
PD 02-MAR-2000.
XX
PF 20-AUG-1999; 99MO-US19361.
XX
PR 21-AUG-1998; 98US-0097550.
PR 12-JAN-1999; 99US-0115639.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Hillman JL, Yue H, Tang YT, Corley NC, Guegler KJ, Gorgone GA;
PI Patterson C, Baughn MR, Lai P, Bandman O, Reddy R, Azimzai Y;
PI Shih LL, Yang J, Lu DM;
XX
DR MPI: 2000-237651/20.
DR P-PSDB: AAY70234.
XX
PT Human RNA-associated proteins useful in diagnosing, treating and
PT preventing cell proliferative, autoimmune, inflammatory and infectious
PT disorders -
XX
PS Claim 9; page 117; 123pp; English.
XX
CC The present sequence is the cDNA encoding human RNA-associated protein-15
CC (RNAAP-15), identified in incyte clone 1879133, derived from LEUKNOT03
CC library. It is expressed in reproductive, dermatologic, gastrointestinal,
CC musculoskeletal and hematopoietic/immune tissues. It has cytosolic,
CC immunosuppressive, antiinflammatory, antiarteriosclerotic, hepatotropic,
CC keratolytic, neuroprotective, antiposoriatic, anti-HIV, antiallergic,
CC antirheumatic, viticude, antiarthritic, ophthalmological and antimicrobial
CC activity. RNAAP antibodies are useful for diagnosis of diseases
CC associated with altered expression or activity of RNAAP. It is used to
CC treat cell proliferative, autoimmune, inflammatory and infectious
CC disorders, like actinic keratosis, bursitis, arteriosclerosis,
CC artherosclerosis, cirrhosis, hepatitis, myelofibrosis, mixed connective
CC tissue disease (MCTD), psoriasis, primary thrombocytopenia and cancer,
CC HIV, allergies, rheumatoid arthritis, uveitis, Crohn's disease, and
CC bacterial, viral and parasitic infections.
XX
SQ Sequence 680 BP; 228 A; 120 C; 134 G; 198 T; 0 other;

```

alignment_scores:

Quality: 451.00 Length: 92
Ratio: 5.184 Gaps: 0
Percent Similarity: 94.565 Percent Identity: 92.391

alignment_block:

US-09-811-045A-1 x AA251264 ..

Align seg 1/1 to: AA251264 from: 1 to: 680

```

1 MetMetGluValLysAspProAsnMetLysGlyAlaMetLeuThrAsnTh 17
  |||
312 ATGATGGAAGTGAAGATCTTAATGAAAGGTGCAATGCTTACCAACAC 361
  |||
17 rGlyLysrTyraIleProThrIleAspAlaGluAlaTyraIleGlyL 34
  |||
362 TCGAAAATGTCATACCAATCACTATGATGACAGACATATGCAATTGGGA 411
  |||
34 yALysGluLysProPheLeuProGluGluProSerSerSerSerGlu 50
  |||
412 AGAAAGAGAAACCTCCCTCTTACAGAGAGACCAATCTTCTTCAGAA 461
  |||
51 GluAspAspProIleProAlaGluLeuLeuCysLeuIleCysLysAsp 67
  |||
463 GAGATGATCCCTATCCACATGATGATGTCCTCATCTGCAAGATAT 511
  |||
67 eMethrAspAlaValIleProCysCysGlyAsnSerSerCysAspG 84
  |||
512 TATGACTGATGCTGTGTCATTCCTGCTGTGGAACAGTACTGTGATG 561
  |||
84 LuCysIleArgThrThrLeuLeuGlu 92
  |||
562 AATGTAAGAAAGTGCATCTTGAA 587
  |||
seq_name: /SIS1/9cgcdata/geneseq/geneseq-emb1/NA2001A.DAT:AAF28296
seq_documentation_block:
ID AAF28296 standard; RNA; 2565 BP.
XX
AC AAF28296;
XX
DT 05-APR-2001 (first entry)
XX
DE LPL mRNA.
XX
KW Lipoprotein lipase; LPL; S447X; cardiovascular; diabetes; obesity;
  deficiency; ss.
XX
OS Homo sapiens.
XX
PN W0200100220-A2.
XX
PD 04-JAN-2001.
XX
PE 23-JUN-2000; 2000MO-CA00762.
XX
PR 24-JUN-1999; 99EP-0202048.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
PA (AMST-) AMSTERDAM MOLECULAR THERAPEUTICS BV.
PA (UYAM-) UNIV AMSTERDAM ACADEMIC HOSPITAL.
XX
PI Hayden MR, Kastelein JJP;
XX
DR WPI; 2001-112388/12.
XX
PT Use of lipoprotein lipase S447X therapeutic for the preparation of a
  pharmaceutical composition for the treatment of coronary heart disease,
  atherosclerosis, angina pectoris, hypertension, diabetes, cachexia and
  obesity
PT
XX
PS Claim 5; Fig 4; S1pp; English.
XX
CC The present invention relates to use of a lipoprotein lipase (LPL)
  S447X therapeutic for the preparation of a pharmaceutical
  for the treatment of an LPL-responsive condition. Diseases that may be
  treated are complete LPL deficiency, chylomicronemia, hyperlipidemia,
  partial LPL deficiency, pancreatitis, hypertriglyceridemia,
  hypobetalipoproteinemia (low high density lipoprotein-cholesterol),
  cardiovascular disease, coronary heart disease, coronary artery
  disease, atherosclerosis, angina pectoris, hypertension,
  cerebrovascular disease, coronary restenosis, peripheral vascular
  disease, diabetes, cachexia or obesity.

```

```

XX
SQ Sequence 2565 BP; 1020 A; 739 C; 806 G; 0 U; 0 other;

alignment_scores:
  Quality: 413.50      Length: 937
  Ratio: 0.884        Gaps: 50
  Percent Similarity: 49.947      Percent Identity: 25.080

alignment_block:
US-09-811-045A-1 x AAF28296 ..

Align seg 1/1 to: AAF28296 from: 1 to: 2565

517 ProHisProGluGluAlaGluAlaArgSerAlaMetIleValHisMe 533
  2 CCCCCCCCCCAAGGAAGC.....
533 rProAspLeuMetAspIleAlaHisAlaArgSerArgSerProProTyra 550
  23 .....GCCACACAGCGCCGCCACCCCAAA 47
550 rGArgTyraArgSerArgSerArgSerProProGluPheArgGlyGln.. 565
  |||
48 GGGCGACGCCAGCGCCAAACCGCGGCCAGCCCCCAGCCCCCGCGAGCC 97
566 .....SerProThrLysArgAsnValProArgGluG 576
98 GGGCACAGCGCGCGCGCGCCGCCAGAGGAGCGCGCCGAGAGAGAG 147
576 LuLysGluArgGluTyrrPheAsnArgTyraArgGluValProProProTy 592
  |||
148 AGCAAAAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG... 194
593 AspIleLysAlaIleTyrrGlyArgSerValAspPheArgAspProPheG 609
195 .....CGGAGCGCGCGCGCGCGCGCGCAAAAGA..... 221
609 uLysGluArgTyraArgGluTyrrPheGluArgLysTyraArgGluTyrrGlu 626
  |||
222 .AGAGACGACAC.....GAAAGAAAGCCCAAGAGACCCGGAAGACA 261
626 ySTyrrTyrrLysGlyTyraIleValIleGlnProArgProSerAlaAsn 642
262 CAGCGAGACAGCCACCCAGCGAGAGACAGCGCGCGCGCGCGAGACA 311
643 .....ArgGluAspPheSerProGluArgLeu 652
312 ACAGCAGCAAAACCCAGGACCGCGCGAGACGAGAGAGAGAGAGG 361
652 uProLeuAsnIleArgAsnSerProPheThrArgGlyArgArgGluAsp 669
  |||
362 GCCAAAGCGCGCGCGCGCGAGAGAGAGAAACCGCAAGAGAGGAGAG 411
669 yTAla.....AlaGlyGlnSerHisArgAsnArgAsnLeu 680
  |||
412 GCGCAGCGCGCGCGAGCAACCCAGCGCGCGCGAGACCAACAGGGGGACA 461
681 Gly.....GlyAsnTyrrProGluLysLeuSerThrArgAspSerHis 695
  |||
462 GGAGGGCGCGCGAGCAACGAGAGAGAGAGAACACCCCGAGACACCGCG 511
695 nAlaLysAspAsnProLysSerLysGluLysGluSerGluAsnValPro 711
  :
512 GGAACAGCGCGAGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 549
712 GlyAspGlyLysGlyAsnLysHisLysLysHisArgLysArgArgAsnG 728
550 .....AAAGAAAGCAACAGAAAGCGCGCGCA 575
728 uGluLysGlyGluGluSerGluSerPheLeuAsnProGluLeuGluT 745
  :
576 CCAGCGAGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 625

```

```

745 hrSerArgLysCysArgGlySerSerGlyIleAspGluThrLysThrAsp 761
    ||| |||
626 CGCACACACACACACAGAGGCCCGCGGAAGCAGAAACGAGAA..... 669
762 ThrLeuPheValLeuPro..SerArgAspAspAlaThrProValArgAsp 778
    ||| ||| ||| |||
670 .....CCAGGGGCAAGCAAC.....CCGAGAGAGAGACC 698
778 LuProMetAspAlaGluSerIleThrPheLysSerValSerAspLysAsp 794
    :||| :||| :||| :|||
699 AGCCGAGAGAACAGAGAGAG.....CACCGCGAGCAGAGAGAGAGA 739
795 LysArg.....GluLysAspLysProLysValLysSerAspLysThrLys 809
    :||| :||| :||| :|||
740 CGGAGAGAGCAGCAGCAGAGAGCCCGCAGAGCCACACCCAGACC 789
790 GAGAGAAAGAAACAGAGAGCCACAGGCGAGCCA.....AGAGAGC 830
809 SarGlySerAspGlySerAlaThrAlaLysLysAspAsnValLeuLysP 826
    :||| :||| :||| :|||
826 roSerLysGlyProGlnGluLysValAspGlyAspArgGluLysSerPro 842
    ||| ||| ||| ||| |||
831 CGAGAAAGGGCGCGGAGAGAAAGACCGCGCAACACGCGGAGAGACAA 880
843 ArgSerGluProLeuLysLysAlaLysGluGluAlaThrLysIleAs 859
    :||| :||| :||| :|||
881 AAAGCGAGAGCA.....AAAGAGCAGCAAAAGACCGAAGACCGCCAGA 924
859 pSerValLysProSerSerSer..... 866
    :||| :||| :||| :|||
925 GCCCACAAACCCACAGAAAGACAGCGGAGAGAGAAACCCACCA 974
867 .....SerGlnLysAspGluLysValThrGlyThrProArg 878
    :||| :||| :||| :|||
975 ACAGGCCGAGACGAGCAGCGGCGGAGAGAGAACCCACCGCCG 1024
879 LysAlaHisSerLysSerAlaLysAspThrArgArgIleSerGlnProAr 895
    ||| ||| ||| ||| |||
1025 AAGCCACAAAGAACCCACCAACACAGAGGAGAGAGAAACACCCAGG 1074
895 gThrArgArgSerLysArgThrValProLysThrSerSerGlnLysSerg 912
    :||| :||| :||| :|||
1075 AAGCCAAAG...AAGAGGACA.....ACAGCGGAGAGCGGGGAGC 1112
912 LnPProValAlaGThrArgArgProArgSerLeuArgLysIleAsnTyrLeu 928
    ||| ||| ||| ||| |||
1113 AGCCCGGCCCGCACAGAAAGACAGAAAGACAGAGAGACAGAAAGAG 1162
929 IleAlaArgGluLysAsnGlu..ArgGluLysArgLysLysSerValAspL 945
    ||| ||| ||| ||| |||
1163 GACCGCAGGAGAAAGCCAGAGAAAGAAAGAGACCGCGGAGAGAAAGC 1212
945 yAspPheGluSerSerMetLysIleSerLysValGluGlyThrGlu 961
    ||| :||| :||| :|||
1213 CAGAG.....AAGCCGAAAGAGAGCAGGAGAAAGCGGCGAGACA 1250
962 IleValLysProSerProLysArgLysMetGluLysPylAspValGluLysLe 978
    ||| :||| :||| :|||
1251 CAGAACAAAGACGCGAGAGAGAAAGAGAGAGAGGA.....AGAAC 1294
978 uGluArgThrProGluLysAspLysIleAlaSerSerThrThrProAla 995
    :||| :||| :||| :|||
1295 ACAAAACACACCGAGGGGCGGAGAGACGAAAGCCAGCCACCGC 1344
995 yLysLysIleLysLeuAsnArgGluThrGlyLys..LysIleGlyAsnAla 1010
    :||| :||| :||| :|||
1345 AGAAG.....GAGACAGCCAGCAACCAAAAGGCGAAC... 1376
1011 GluAsnAlaSerThrThrLysGluProSerGluLysLeuGluLysThrSe 1027
    :||| :||| :||| :|||
1377 AAAGGGAGAGGCGCAAAAGCAGACCCAGCAAGAAAG.....ACAGG 1417

```

```

1027 rSerLysIleLysGlnGluLysValLysGlyLysAlaLysArgLysVala 1044
    :||| :||| :||| :|||
1418 GACAGAAAGAGCGCGCCGAGAAAGAAAGAAAGCA...GCCACAGAAAGAG 1464
1044 IaGlySerGlnGlySerSerSerThrLeuValAspTyrThrSerThrSer 1060
    ||| :||| :||| :|||
1465 CCCAGGAGGAGGCGCCAGGCAAG.....AGAACCCCAAGAGAA 1505
1061 SerThrGlyLysSerProValArgLysSerGluGluLysThrAspThrLys 1077
    ||| ||| ||| ||| |||
1506 CGGACCGGAGAGAGCCCAACCAAGCCCAAGAA.....ACAGA 1546
1077 SarGThrValIleLysThrMetGluGluTyrAsnAsnAspAsnThrAlaP 1094
    :||| :||| :||| :|||
1547 AAAGCCCGCGCAGAGACA.....GACCCACACACAGC 1578
1094 roAlaGluAspValIleIleMetIleGlnValProGlnSerLysTyrPasp 1110
    :||| :||| :||| :|||
1579 AGAGAGAGAG.....AAGACCCCGCAGAGAGGCGCCAGAC 1613
1111 .....LysAspAspPheGluSerGluGluLysValLysThr.... 1123
    :||| :||| :||| :|||
1614 GCAACCGAGAACAGAACGAGAGGAGCAGAAAGCGCGAAGACACAGAG 1663
1124 .ThrGlnProIleGln..SerValGlyLysProSerSerIleIleLysAsn 1139
    ||| ||| ||| ||| |||
1664 GAACACAGCCACACAGCCCGCACACACAGACAAAGACAGAGAAAGAAACA 1713
1140 ValThrThrLysProSerAlaThr.....AlaLys..... 1149
    ||| ||| ||| ||| |||
1714 GACAAACAAAGGACAGACGACACCCCGCCCGCCCAAGAAAGAAAGGCC 1763
1150 .....TyrThrGluLysGluSerGluGlnProG 1159
    :||| :||| :||| :|||
1764 GCGAGGGAGAAAGAGCCGAACCCAGCACACAAAGAGGCGAGAAAGAAA 1813
1159 LuLysLeuGlnLysLeuProLysGluLysSerHisGluLeuMetGlnHis 1175
    :||| :||| :||| :|||
1814 AAAGAACCGGAAAGAACCCCAAGACACAAAGAAACACAGACAGACCA 1863
1176 Glu.....LeuArgSerSerLysGlySerAlaSerSerGluLysGln 1189
    ||| :||| :||| :|||
1864 GAGGACACCGGAGCAGACGACCCAGCGCGCGCGAGCAAGAAAGAGG 1913
1189 Y.....ArgAlaLysAspArgGluHisSerGlySerg 1200
    :||| :||| :||| :|||
1914 AGCAAAACAGAGAGACCAAGAACAGACCGACGCAAGAAAGAGGAGAGA 1963
1200 LuLysAspAsnProAspLysArgLysSerGlyAlaGlnProAspLysGlu 1216
    :||| :||| :||| :|||
1964 GCGCGGGGGGACCAAGCAGCGGAAACACCGCGGCAAGCCAGAGAGAC 2013
1217 SerThrValAspArgLeuSerGlnGlnGlyHisPheLysThrLeuSergI 1233
    :||| :||| :||| :|||
2014 CAGAGGGGAGAGAACACACAGAGAGAGAAACAAAGAAACAGACCCACA 2063
1233 nSerSerLysGluThrArgThrSerGluLysHisGluSerValArgLys 1250
    ||| :||| :||| :|||
2064 G.....AAAGCCACGAGAAAGAAAGAAAGAGCAAGCAAC...AGAAGCA 2104
1250 erSerAsnLysAspPheThrProGlyArgAspLysLysValAspTyrPasp 1266
    :||| :||| :||| :|||
2105 AGCGGAGAACCCGAC.....GGAAGAGAGAGAGAGAGAGAGAG 2139
1267 SerArgAspTyrSerSerSerLysArg.....GluArgGlyG 1281
    :||| :||| :||| :|||
2140 AGAAAGCCGAGGAGGCGCCAGCAACGAGAGAGAGAGAGAGAGAGAG 2189
1281 LuLeuAlaArgArg.....LysAspSerProProArgGlyLys 1293
    ||| ||| ||| ||| |||
2190 AGGAACAGCGGAGAGAAAGCCGAGAGCCAGAACCGACCGCGAGAGAGA 2239
1294 GluSerLeuSerGlyGlnLysSerLysLeuArgGluGluArgAspLeuPr 1310

```



```
289 .....ProTrpAsnSerIlePheV 295
4278 GAGGAGAGTACTGATTCATCGAAAAGAGAGATGGAGCATATACCGA 4327
295 aIcIyInSerIleuHisIcIyInIleuPro..... 306
4328 AGATTCGATCGTTGACAGAGAGACGATGACCAAGAAAGAGAGATCC 4377
307 ...ThThGlyProValArgIleAsnAlaIleArgProIyGlyIylar 332
4378 AGAAGACGACGCTGACTCGCAGACATCGAAGAAAGAAAGAGAG 4427
322 gProGlyTrpGluHisSerAsnIylSerIyTrleuValSerProProg 339
4428 GCCGAG..... 4434
339 IngIleArgArgIyIuArgSerCysTrArg.....Ser 351
4435 .....AGGAGATCGAGCGCGCATGCGAAGGTAAAGCCGCTGAGC 4476
352 IleAsnArgIylArgHisHisSerIuArgSerGlnArgThrcInSerPr 368
4477 GTAAACGCGGACTAGAGGGTGCCTCTCGAGCGAGATGAGTGGAACTG 4526
368 oSerIleuProAlaThr.....ProCysPheValProV 379
4527 GACGTGACCGCGAGACGACAAAGATMAAGCTGAGTGCAGATATCAT 4576
379 aIProProProLeuTyrrProProProHisThrcInProleuPro 395
4577 TGCCACTGCGCAAGGATATTG.....CCAAAGTCACCA...CCGAGAGC 4617
396 ProGlyValProProGlnPheSerProGlnPheProSerSerGlnPr 412
4618 AGTGGCGGAAGCCCGCGAGAGAGAGCTCTCAT.....CGCCAAAC 4658
412 oProThrIaGly.....TyrS 418
4659 CCGAGAGAGAGCTAAGCAAAAGACTTGGATAGACAGAGCATCAAT 4708
418 eValProProProGlyPheProProAlaProAlaAsnIleSerThra 434
4709 CGAGGCCCGAGTCGACGTTCTACCG.....TCA 4737
435 CysPheSerProGlyValProThrIaHisSerAsnThrcProthrt 451
4738 ATCTTAAGAAAGCTAACCG...AAGAGAGTTCTCGGCTACCATTTAG 4784
451 rGlnAlaProLeuLeuSerArgIuGluIuPheTy..... 462
4785 TCGGGGGCCACCACTGCTCGACAC.....TACCGAGGATGAGCGCAT 4828
463 .....ArgIuGluAsn..AspIylsIylArgIuSerIy 473
4829 TCCGCTGATTCAGATTAAGGAGACCTCGTATCGAGAGAAATATGTGA 4878
473 sPheProTyrrSerGlySerSerTyrrSerIuArgSerTyrr...ThAsps 489
4879 AAGAGAGACCAAGGAAAGCGAGCGCATGTAGTGGCCACCGTACGACT 4928
489 eSerGln.....GlyLeuAlaGlnHis..IleHisAlaLeuTh 501
4929 CTTCGCGAAGCAGCTCCATTTGGCAATCGATACCATCTTCCCTACACTAC 4978
501 rIeuSerProSerAlaAlaHisThrcInAspIleuHisAspHisProH 518
4979 CAAGGATGCGCCCAAGAGCGCAACGCTTAGCGGAGACTCCCG 5028
518 IsrProGluGluAla.....GluAlaArgSer 527
5029 ACTCAGGCGAAGGGTGGCTGCCATGAGATGACATTTGAGGCGCAGAG 5078
528 AlaMet.....IleValHisMetProAspIleuMetAspIleuHis 541
5079 AACCTTCTCCGCGATGTCAATCAAAAGCCGTGAC.....GAATGTGACGA 5122
541 sAla..ArgSerArGserProProTyrrArgTrArgTrArGserArGser... 556
5123 TTTAACCGTTCATAGAGAAAGCCGATCTTGGCTATACGAAAGAGACAGTC 5172
557 .....ArgSerProProGluPhe 562
5173 AGAGCGCGCGCGCAAAAGTATCTCCATTAAGAGAGAGTCCGCAAGAA 5222
562 eArgGlyGlnSerProThrIylsArgAsnValProArgGluGluIylsGlu 579
5223 GAAAGAGAGCGAGACGACGAAAGAGAAATA...AGGTAGAGATGATCCA 5269
579 rGluTyrrPheAsnArgTyrrArgIuValProProProTyrrAspIleIys 595
5270 ATTGGGGGATGAACCGAATTAAG.....TGCACATCTCTC 5304
596 AlaTyrrTyrrGlyArgSerValAspPheArgAspProPheGluIylsGlu 612
5305 ACCTACTACTCAAGAGT.....CG 5324
612 gTyrrArgIuTrpGluArgIylsTyrrArgIuTrp.....T 624
5325 GTACAGAGAGTGGCCGAGAAAG.....TCTGTATATGAACTACAGT 5368
624 yrlGluIylsTyrrIylsGlyTyrrAlaValaIylacInProArgProSer 640
5369 GCGAAGAAAGACAGAGAGATGTGTGAGGCCACTACAGTCAATTAACCAAG 5418
641 AlaAsnArgIu..... 644
5419 AAAACAGGAAGATCTAATGAGACAGGTTAAGATTAAGAGAGCATGA 5468
645 .....AspPheSerProGluArgIleu..... 652
5469 CAAAAATAGAGTCTGGAATCATTTACGAAAGAGCAAGCTTAAGAGCGC 5518
652 .....P 653
5519 TACACTCCCGAGAAAGAGAGACAGACATTTACATCGATGAGAGC 5568
652 ..... 652
5569 TTCCTCCCAACTGGCGAGCCGACCACTGTGCCACCAATACGCCAA 5618
653 .....P 653
5619 GACCGGAGATACCGGCTCCATTTGAAGTCCCTACCATTTGAGGAGGC 5668
653 rOleuAsnIleArgAsnSerProPheThrArgIylArgArgIuAspTyrr 669
5669 CATTTAGGTTGAGGTGACGCGCAAGCAAGAAAGCCCAAAACGAGTCCAG 5718
670 AlaAlaGlyInSerHisArg.....AsnArgAsnIleuIylGlyAsnTy 684
5719 CTCGAGAGAGGCTATCAAGACGAAAGTGCACCTTGCTTGAAGGAG 5768
684 rProGluIylsLeuSerThrArgAspSerHisAsnAlaIylsAsp..... 698
5769 ACCTCCGACCTGAATCTGTGACAGAGAGAGCGTCAAGGAACACGTGACGA 5818
699 .....AsnPro..LysSer.....L 703
5819 GACGAATATCCAAAGAAATCCCACTACCTCCGACCTGAATCTGAGGAGA 5868
703 yslGluIylsGluSer.....GluAsn 709
5869 AAGATTAAGAGTCTCCATTCGCTTACAGAGAGAGCTTGTGACCGGAATCT 5918
710 ValProGlyAspGlyIylsGlyAsnIylsHisIylsArgIylsArgTr 726
```



```

1184 ..... 1184
7730 AAAATCTCTTGGCCCTCCAGGAAGCTTCAGACCAGCTCCGTTGCAG 7779
1185 ..... SerSerGluValArgAlaLysAspArgLHis 1196
7780 AAAGATTAAAGACGACAGCTAGGAAGTGTAAAGAAAGAACTAGAGCAGAG 7829
1197 SerGlySerGlu..... 1200
7830 TCGGTGGCGGAAAAATCTCTTGGCTCCAGGAAGCTTCAGACCAGC 7879
1201 ..... LysAspAsnProAspLys..... 1206
7880 ATCCGTCGACGAAGATATTAAAGACGAAGCTAGAAAGCTAAGGAAGAAAT 7929
1207 ..... ArgLysSerGlyAlaGln..... ProAspLysGluSer 1217
7930 CTAGCCGAGAGTGGCGGCCGGAATAATCTCTTGGCTCCAGGAAGAAAGCT 7979
1218 Thr..... ValAspArgLe 1222
7980 TCCAGACACAGCTCCGTTGCAGAAAGTGTAAAGATGAAGACCTGATTAAGTC 8029
1222 uSerGluGlnGlyHisPheLysThrLeuSerGlnSerLysGlu.... 1237
8030 CAAGAAGAAATCTAGCGCGGAATCGATGGCCGAAGACGGTAAGCCGACAG 8079
1238 ..... ThrArgThrSerGlu 1242
8080 GTATCAAGAGGTGATCATATCTCTCTGAAGGAAGTTTACAGACCAAGAGTCT 8129
1243 LysHisGluSerValArgGlySerSerAsnLysAspPheThrProGlyArg 1259
8130 GTGGCGGAGAGGTGTTAAAGACATCTGTCAAAATCCAGGAACCATCCCGC 8179
1259 GAspLysLysValAspTyrAspSerArgAspLysSerSerSerLysArgAla 1276
8180 TCGAAGAAATCTGTA..... GCAGGAAGTGTACAGCGGGAATAGTCTTA 8220
1276 rGAspGluArgGlyLysLeuAlaArgArgLysAspSerProProArgGly 1292
8221 GAGACGACCAATCTCTCTTGGAAATCCAGGAGCGCTCAAGGCCA..... 8264
1293 LysGluSerLeuSerGlyLysLysSerLysLeuArgGluGluArgAspLe 1309
8265 ... GAATCAGTC..... GTTGACACGCTTAAGACGAACTGTAAGAA 8302
1309 uProLysLysGlyAlaGluSerLysLysSerAsnSerSerProProArgAla 1326
8303 GCAGAAAGATCGTCCGCAATTAAGAAACAGAGGTGTATTAACCTCCAAAG 8352
1326 sPLysLysProHisAspHisLysAlaProTyrGluThrLysAspProCys 1342
8353 CTAAAG..... CATGATAAGTCTCCAAAGGAATTTTACAAACCATG 8393
1343 GluGluThrLysProValAspLysAsnSerGlyLysGluArgGluLysHis 1359
8394 TCTATGACGAGACATATTATGGGAAGACGCCGACCAACCATTAAGGCATC 8443
1359 sAlaAlaGluAlaArgAsnGlyLysGluSerLysSerGlyAlaAsnCysHis 1375
8444 ACAACACAGAAAGCTCGAGGAATCGATTCTGAAACCATTAAGGCCCTTCA 8493
1376 ValTyrLeuThrArgGlnThrLeuProTyrArgSerThrPheLeuGlu 1392
8494 GTCTTAGGAGCAGAGA... AGTCTCAATTGGCTTGAAGAAAGCCTTAGG 8540
1392 YArg 1393
1:1:
8541 CCAG 8544
seq_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT.AAH81795

```

```

seq_documentation_block:
ID   AAH81795 standard; DNA; 9027 BP.
XX
AC   AAH81795:
XX      21-SEP-2001 (first entry)
XX
DE   Human differential transcription-associated cDNA seq ID 304.
XX
KM   Differential transcription; human; rat; tumour cell; cytostatic;
KW   Ras modulator; Class II tumour suppressor gene; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   W0200157058-A2.
PD
PF   09-AUG-2001.
PI
PR   31-JAN-2001; 2001WO-EPO1003.
PP
PX   31-JAN-2000; 2000DE-1004102.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A, Hinzmann B, Schaefer R, Zuber J, Tchernitsa O;
PI   Grips M, Hellriegel M, Schmitz A, Sers C;
DR   WPI; 2001-483415/52.
XX
PT   Nucleic acids differentially expressed between tumor and normal cells,
PT   useful for diagnosis or therapy of tumors and for screening active
PT   agents -
PS
SS   Disclosure; Page 457-459; 579pp; German.
XX
CC   This invention describes a nucleic acid (I) with differential expression
CC   between tumour and normal cells and which has cytostatic activity. (I)
CC   work as modulators of Ras activily by inducing expression of tumour
CC   suppressor genes. (I), and polypeptides encoded by them, are useful as
CC   targets for diagnosis or therapy and in screening to determine the
CC   effects of an active compound (potential pharmaceutical) on a cell line,
CC   particularly for diagnosis and treatment of tumors, especially by
CC   modulating expression of (I) (by gene therapy, antisense RNA or ribozyme
CC   methods) or by modulating the amount and/or location of (I)-encoded
CC   peptides (by administration of the polypeptide or its activator,
CC   antibody (optionally as a conjugate) or inhibitor). The method allows
CC   identification of many Class II tumour suppressor genes (i.e. genes that
CC   are not primary targets for tumour-initiating mutations).
CC   AAH81492-AAH82316 represent the human and rat derived nucleic acid
CC   fragments described in the method of the invention.
XX
SQ   Sequence 9027 BP; 2206 A; 2831 C; 2223 G; 1767 T; 0 other;
XX
alignment_scores:
    Quality: 400.50          Length: 1784
    Ratio: 0.515            Gaps: 77
    Percent Similarity: 43.610     Percent Identity: 20.684
alignment_block:
US-09-811-045A-1 x AAH81795 ..
Align seg 1/1 to: AAH81795 from: 1 to: 9027
31 AlalaleglylsglyslglsyPProPhleuProgiungluProSerSe 47
||||| ||||| :||| |::||| ::||| :||| :|||
36f GCCCTGGTGAAGCGGCGCTAAATCCTGCATCCTCGACGACGAGGCCAACG 416
47 rserterglutguaspaprolleproAlaglutenuleucysLeutiEc 64
:::||||| | |||||
417 GGCG.....GTGAGCTGGAGTGCCGTCGAGC 442

```

```

64 yslysaPlleMeThrAspAlaValAlleProCysCysGlyAsnSer 80
   :::::::::::::::
443 TGAGGAGATGATGAAAG.....CAG 465
81 SerCysAspGlu.....CysLeuArgThr 89
   :::::::::::::::
466 GGCTAGAGAGAACGCAATTCAGAAAAAGTGGCAGCCTTCGACTCAT 515
89 rleuValuGlnSerAspLysHisThrCysProThrCysHisGlnAsnAspY 106
   |||||
516 GTTGTGGAG.....AAGATG 532
106 alSerProAspAlaLeuIleAlaAsnLysPheLeuArgGlnAlaVal... 121
   ||::::|
533 TGAACCTGGGGCAGAGAGAGACCCAGGCCAGCGCGTCACG 582
122 .....AsnAsnPheLysAsnGlnThrGlyTy 130
   ||::::|
583 GAGACTCACCACTTGGCAGATTAAATGAGAGAGAAATGAA..... 624
130 rThrLysArgLeuArg..... 135
625 .....AGACTCGTGCTGCTTGGCATCAGTATTCTAGTAGATG 667
136 ..... 139
668 GCAGCTCTTTGATCTCAGCGTCGTCGCCAGAACGTAACAACCAAGCT 717
140 ProPheLeuPheLeuValProProProArgProLeuSerGlnArgAsnLe 156
   |||
718 CCT.....GAGCCTCCCAACCTTACAGCCTTGTTCCGGA 752
156 uGlnProArgSerArgSerProIleLeuArgGlnGlnAspProValValP 173
   ::::|
753 GTCAGCAAGTCTCGTCACCAACCCCAAGACAGAGAAG..... 795
173 heArgTyThrValSerProThrCysSerAspThrLysThrAlaGlySer 189
796 .....AAAAAGAAAGATAGAGAGCGC 819
190 CysSerAspSerGlyThr..... 195
820 AGGTCAAGAGACAGCTCTCTCGACGGAGAGAAAGAAAGCTCAAGAA 869
196 .....LeuSerArgLeuP 200
870 GAAGAGACAGGTCAAGATCTGAGTCCAGAAAGTAAGCATAGTCTC 919
200 roAlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAlaPro 216
   ||::::|
920 CCACCTCCAAAG.....AGCAACGTAATCTAAGGACAAA 954
217 ProValSerGlyAsnProSerSerAlaProAlaProValProAspIleTh 233
   ::::::|
955 AAGCGAAAGCGGTCTCGAAGTACACACACACCC..... 990
233 rAlaThrValSerIleSerValHisSerGlnLysSerAspGlyProPhea 250
   ||::::|
991 .....AAGAGCCGGCGGCCACACCTTCAACTCTGCTGACTCTGCTT 1033
250 rGAspSerAspAsnLysLeuLeuProAlaAlaAlaLeuThrSerGlnHis 266
   ::::::|
1034 CCTCTCCGAT.....ACTTCCCGC 1053
267 SerLysGlyAlaSerSerIleAlaIleThrAlaLeuMetGlnLysGly 283
   ||::::|
1054 AGTCGGTCTCGAAGTCTGCAAGCTAAACT.....CATACACTGC 1094
283 yAlaProGlyThrSerProThrAsnSerIlePheValGlyGlnSerLeuL 300
   ::::|
1095 CTGGCTGGGGAAGTCT..... 1113
300 euHisGlyGlnLeuIleProThrThrGlyProValArgIleAsnAlaAla 316

```

```

1114 .....TCCCTGCTCAGCG.....CGA 1131
317 ArgProGlyGlyLysArgProGlyTyThrGlnHisSerAsnLysLeuGlyTy 333
   |||
1132 CGGGGGAGGAGATCGGCTTCACTGAACCAAGTACTACAGCAGCACA 1181
333 rleuValSerProProGlnGlnIleArgArgGlyLysValArgSerCysTyra 350
   |||
1182 ACGGCTAGTACCGGAGACTGCTACGAAACGCTTACAGCCCTTAAG 1231
350 rSerIleAsnArgGlyArgHisHisSerGlnLysSerGlnArgGln 366
1232 AAGACAAAGATTAAGACAGAGAGAGAAATCTGCACTGACCTAGCCCC 1281
367 SerProSerLeuProAlaThrProCysPheValProValProProProP 383
   ||::::|
1282 TCTCCGAAAGAGACAGACAGGCCAGAACCCCTGCTCCCTCGCT 1331
383 oleu.....TyrrProProProHisThrLeuProL 394
   |||
1332 CCTTGTGAGCGACATGGCGGCTCCCAACACCTTGCACACCCCT 1381
394 eu.....ProProGlyValProProProGlnPheSer..... 404
1382 TAGCCAGAGGCCAGTGAACCCCACTGAGGCTCTCACTCGGAGAC 1431
405 .....ProGlnPheProSerSerGlnProProThrAlaGlyTyse 418
   ||::::|
1432 CGTTCAACCACTTAAGTCTCCGAGAAACCTTCCAGCTCTTCCACGA 1481
418 rValProProProGlyPheProProAlaProAlaAsnIleSerThr...A 434
   |||
1482 GAGCAAGCCCACTTC.....CCTCACTTACCAAGTTTCTCGGATG 1525
434 lAcysPheSerProGlyValPro..... 441
1526 CAGGCTTTCCTCCAGAAAGTCTTAACCTGCTCAGCTCCAGGGTCCAC 1575
441 ..... 441
1576 CGAGAGATTCTTCTCTCCACATCTAAGATCTCACATGGCCGAGC 1625
442 .....ThrAlaHisSerAsnThr..... 447
1626 AAAAGGAGTAATACATTTCTCATACCCCTCCGATAGTGGGAGGT 1675
448 .....MetPro 449
1676 CCCGTAGCCCTGCCACCGCTAAGAGAGGCGATCTCGTTCGAACCCCT 1725
450 ThrThrGlnAlaProLeuLeuSerArgGlnGlnPheTyrrArgGlnAs 466
   |||
1726 ACCAAGAGAGTCAATTCTGATCCCGATCTCCAGTGGCGTAGGTCAAG 1775
466 nasPlys.....GlyArgGlnSerLysPheProTyrrSerGlySers 480
   |||
1776 GTTGCACAGAGGTGGGAGATCTAAGAACCCCGCAGAGCTGGCGGCT 1825
480 er.....TyrrSerArgSerSerTyrrThAsp..Se 489
   |||
1826 CTAGGTCTCTCAGGACAGCAGGCTGTGTAGAGCAGAAATATCCAGAGA 1875
489 rSerGlnGlyLeuAlaGlnHisIleHis..AlaLeuThrLeuSerPro. 504
   ::::|
1876 AGAGCAGAGTCTAGTCTAGCAAGGCGAGGAGTCCACTTAATATCCC 1925
505 .....SerAlaAla 507
1926 AGCCACTAGGGGTAGATCTCTTCTAGAACACAGCCCGGGGAGAGGT 1975
508 HisThrLeuAspLeuLeuHisAspHisProHisProGlnGlnAlaGly 524
   ::::::|

```



```
986 silealaserthrthrproalalyslylleysleuan..... 1000
|||||:|||||:|||||:|||||:|||||:
3751 ....GCTACTGCATCACCTCCTCTAGACAGAAAGAAATTTAGTCCCTTTC 3796
1001 .....Arg 1001
3797 CAGTACAGAGATAGCCCTGAGTCTTCACTGTATTCAAGACACACTTGA 3846
1002 gluthrlylslylleglyasnalaasethrthrlysgl 1018
|||||:|||||:|||||:|||||:|||||:
3847 ACCCGCCCAAGAAAGAGTGTCTGTGTCATCTCCAGAAACAAAGA 3896
1018 uProserglulysleugluserthrSerSerLyslleysglnglulysv 1035
|||||:|||||:|||||:|||||:|||||:
3897 GCAAAATATGTCATTGCTCTACGTCACGCAAGATGAAAGATTAAATGAGG 3946
1035 allysgllyalalysarglyvalalaglyserglulysSerSer 1051
|||||:|||||:|||||:|||||:|||||:
3947 TGGTAGAGAGATCTGAAAGAACCCGAGCCAAATCCTGTCATTGTCT 3996
1052 ThrleuValAspTyrThrSerThrSerThrThrThrThrThrThrThr 1068
|||||:|||||:|||||:|||||:|||||:
3997 TCAGAACTTAAGAAATGTCACCAAGTACTTGAATCATCTCCT.... 4041
1068 glySerSerLysLysThrAspThrLysArgThrVallleLysThrMetg 1085
|||||:|||||:|||||:|||||:|||||:
4042 .....GAGTAGAAGAAAGAGCCCTGCTGTCTTGTACTCTTTC 4078
1085 lugalutyrAsnAsnAspAspThrAlaProAlaglualaspvalilleleMet 1101
|||||:|||||:|||||:|||||:
4079 ATCAGAGCCAGTCAACAGCCTTCTTGGAGCA..... 4110
1102 llegluValProglu.....SerLysTrpAspLysAspAspPheglu 1116
|||||:|||||:|||||:|||||:|||||:
4111 GTAGAGAGCCCTTCATAGCTCTTGGGAGGAGCCACATTTTCTCC 4160
1116 rgluglulualspvallysththrthgluProilleglu.....SerValg 1131
|||||:|||||:|||||:|||||:|||||:
4161 AGAACATTAAGAACTG...TCTACTCCCTCAGGAGCAACAGCTTTC 4207
1131 lylsProSerSerilleLysasnValThrThrLysProSerAlaThr 1147
|||||:|||||:|||||:|||||:|||||:
4208 GATCACTT.....TTAGAAATTAGAAACTCAGCCCACTTGGTACA 4248
1148 AlalysTyrThrLysglulysglusergluInProglulys..... 1160
|||||:|||||:|||||:|||||:|||||:
4249 GAAATGATACTGGATTTTCTTCTGAGGTTAAAGAGATTTGAATGAGCC 4298
1161 ....LeuglulysleuProLysglualaserHisgluleuMetgluHisg 1176
|||||:|||||:|||||:|||||:|||||:
4299 GTTCTTAAATCAGCTGGAACAGATCCATCTCTAGACATGAAGAACAAAT 4348
1176 luleuArgSerSerLysLysSerLysSerLysSerLysglulysglulys 1189
|||||:|||||:|||||:|||||:|||||:
4349 CGACACAGATCTCT...GGACACAGCAAGTCTGAGTTATCCCAAGATGCA 4395
1189 ..... 1189
4396 GTGGAAGAGGAGGATGCTTCAATTCAGAGCATCTTTCACCTGTCT 4445
1190 .....ArgAlaLysAspArggluHisSerLysglulys... 1200
|||||:|||||:|||||:|||||:|||||:
4446 TGATGCTGTACCCAGACACCTCGAGAGAAAGATGTTCTGCATCTT 4495
1201 .....LysAsp.....AsnProAspLysArg...Lys 1208
|||||:|||||:|||||:|||||:|||||:
4496 CTCCTGAAGATGAAGATGTTTACCCAGAACTCCATCAAGAGAGCAGG 4545
1209 SerglualaglInPro.....AspLysgluserThrValAspArgle 1222
|||||:|||||:|||||:|||||:|||||:
4546 TCTGGGTCTTCTCCAGAGACTTAGAGATGGTCTGGAGCTCCCTCGAGGCA 4595
1222 uSergluIngluHis.....P 1228
```

```
4596 CAGCCTGTCTGGTCTCTCTCGAATGAAGATATACCTAGAACCCAT 4645
|||||:|||||:|||||:|||||:|||||:
1228 helysThrleuSerLysgluThrThrThrThrThrThrThrThrThrThr 1242
|||||:|||||:|||||:|||||:|||||:|||||:
4646 TTAGAGGAGAGAGCAATGTGATCTTCCCGAAGCAAGAGCTTTCCT 4695
1243 LysHisgluserValArgglySerSer.....As 1252
|||||:|||||:|||||:|||||:|||||:
4696 CAGACTCTGAGCGGAGAGAGTGTTCATCATCATCCAGAGCTCAACAA 4745
1252 nLysAspPheThrProgluYargAspLysValAspTyrAspSerArg 1269
|||||:|||||:|||||:|||||:|||||:
4746 CAATGTCTTACCCCGAGAGAAAGAGGGGTGCAATCATCATCACTTG 4795
1269 spTyr.....Ser 1271
|||||:|||||:|||||:|||||:|||||:
4796 ATCAGAAAAGTGGCTCGAGCTCCCTGGGGCAGAGAGTCTTCGGGA 4845
1272 SerSerLysArgArgAsp.....gluArggl 1280
|||||:|||||:|||||:|||||:|||||:
4846 TCTCTCAGAACTTGATGTGAACCCAGTGCATCCCTCAGAAAGAG 4895
1280 Ygluleu.....AlaArgArgLysAspSerProProArgg 1292
|||||:|||||:|||||:|||||:|||||:
4896 TGACTCAGACTCTTCTCCAGATTCTAAAGCCAGACAGAACCCCACTTC 4945
1292 lylsgluserleuSerLysglu.....GluLysSerLys 1302
|||||:|||||:|||||:|||||:|||||:
4946 GGCAGAGAGTGTCTGTGATCATCTCCAGAGTTGACAGCAATCTCGA 4995
1303 leuArg.....gluLysArgAspLe 1309
|||||:|||||:|||||:|||||:|||||:
4996 CTATCCCTCGGCCAGTAGCTGTGTTCTCCCTGAGAGAAAGATTA 5045
1309 uProLysLysglualagluserLysSerAsnSerSerProProArg 1326
|||||:|||||:|||||:|||||:|||||:
5046 GCCAAGAGCAGCACCCAGGAGCAGAGTGTGTGATTCCTCTCTGAC 5095
1326 sPLysLysProHisAspHisLysAlaProTyrLysThrThrThrThrThr 1342
|||||:|||||:|||||:|||||:|||||:
5096 CTAAAGCTCCAGCCCCCGGCGCCCTTCCAGACAGACAGATCAAGTTCA 5145
1343 GluLys.....ThrLysProValAspLysAsnSerLysgl 1355
|||||:|||||:|||||:|||||:|||||:
5146 TCAAGCAAGGAGCAGAGCCCTTCTCTGAAGAGAGCAGAGTACCGAGTC 5195
1355 uArgglulysHisAlaAlaglualargAsnLysgluserSerLysgl 1372
|||||:|||||:|||||:|||||:|||||:
5196 CTCCTCGAATCAGCGCCCAATCCAGAACTGTGCGAGAGTTCCAGGT 5245
1372 laAsnCyHisValLysThrThrThrThrThrThrThrThrThrThrThr 1388
|||||:|||||:|||||:|||||:|||||:
5246 CATCACAGAGAGCCAGAGCAAGTCTGTACACCACTCCAGCTGCGAGC 5295
seq_name: /SIDSI/gcgsdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA107156
seq_documentation_block:
ID AA107156 standard; DNA; 24387 BP.
XX
AC AA107156;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SFG ID NO: 9844.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
```

PD 02-AUG-2001.
 XX 17-JAN-2001; 2001WO-US01339.
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218299.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231245.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0231415.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0233397.
 PR 14-SEP-2000; 2000US-0233398.
 PR 14-SEP-2000; 2000US-0233399.
 PR 14-SEP-2000; 2000US-0233401.
 PR 14-SEP-2000; 2000US-0233401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-465570/50.
 DX


```

13971 .....GAATGAAGAAATCTCGCATTTCTCAAGGCGGAGAGGTCTC 13929
      ||| ..... ||| ..... |||
623 rPtyrGluSerTyrTyrLysGlyTyrAlaValAlaGlnProArgPro 639
      ||| ..... |||
13928 TC.....TCTTCACACAGGTCC 13912
640 SerAlaAsn.....ArgGluAspPheSerProGluArgLe 651
      ||| ..... |||
13911 AAAGCAAAATCTCGCTTCTTGTGAGGCGACCTTTCAGGGTCTTCC 13862
651 uLeuProLeuAsnIleArgAsnSerProPheThrArg.....GlyA 665
      ||| ..... |||
13861 ATGCCCTAAAGCAAAAGTCACACACACCCAGCGCAGTCGCTGTGAT 13812
665 rGArgGluAspTyrAlaAlaGlnSerHisArgAsnArgAsnLeuGly 681
      ||| ..... |||
13811 COTCCCAAACTAAAGCTAAATCTAGACGCCACCCAGCGCAGTCGCTCC 13762
682 GlyAsnTyrPro.....GluLysLeuSerThrArgAspSe 693
      ||| ..... |||
13761 AGTCTCTCCGCCACCTAAACGAAATCTAAGACACATCAAGCAAG 13712
693 rHisAsnAlaLysAsp...AsnProLysSerLysGluLysGluSerGluA 709
      ||| ..... |||
13711 TCATTCGAGTTTATCTCTCCATCTCTAAAGTGAA.....TCGGGA 13671
709 snValProGluLysAspLysGlyAsnLysHisLysHisArgLysArg 725
      ||| ..... |||
13670 CACCAACCGGAGCAAGGTCATTAACAGTCCCGCAGGCCAATGAGCAATGT 13621
726 ArgAsnGluGluLysGlyLysGluSerGluSerPheLeuAsnProGluLe 742
      ||| ..... |||
13620 GTAAAGCCACAGAGAGAGCTGTTTGATCATCATCACTGACCTGAGT 13571
742 u.....LeuGluThrSerArgLysSerGlySerGlyIleA 756
      ||| ..... |||
13570 GAAATCTAGAGACCCCTTCTAGACATAGCTGCTCAGGGTCTCTCC... 13525
756 spGluThrLysThrAspThrLeuPheValLeuProSerArgAspAspAla 772
      ||| ..... |||
13524 .....CCTAAGCTGAAATCTAGC 13507
773 ThrProValArgAspGluProMetAspAlaGluSer..... 784
      ||| ..... |||
13506 ACACCTCCAGACAGAGCCCATGTATCATCTCCACAAACCAAAAGT 13457
785 .....IleThrPheLysSerValSerAspLysAspLysArgGluL 798
      ||| ..... |||
13456 GAAGGCATTAATATACCAACAGCAAGAACCAATCTGCTCTCTCTC 13407
798 ysAspLysProLysValLysSerAspLysThrLysArgLysSer..... 812
      ||| ..... |||
13406 CAAGTCTAGTAGGGTGAAGTCGAGAAACAATCCACGGCGAAGACATCA 13357
813 .....AspGlu 814
13356 GTATCTCCCTGCTCCAAATGTGAATCCAGATTGTGCCAAGATACAGTCA 13307
814 ySerIleThrAlaLysLysAspAsnValLeuLysProSerLysGlyProG 831
      ||| ..... |||
13306 TTCTGGGCTCTCTCTCAACAGATACCAAAAGTGAACCTGAAACACCGCCA 13257
831 InGluLysValAspGluAspArgGluLysSerProArg..... 843
      ||| ..... |||
13256 GACAAGATCACTCAGGGTCTATTTCACCATACCCCAAGTAAGGCCCA 13207
844 .....SerGluProProLeuLysLysAl 851
13206 ACTCACCGGGCCAAAGTCTTCTGTGATCAAAATCACCATGTCGCCAAGA 13157
851 aLysGluGluAlaThrLysIleAspSerValLysProSer..... 864
      ||| ..... |||

```

```

13156 GAAGTCTAAAGACTCACTAGTTCAAAGTTGCCCTGGATCCCTCTCTCT 13107
865 .....SerSerSerGlnLysAspGluLys.....Val 873
      ||| ..... |||
13106 GTGCAGAGACTAAATCTACACACCCAGCGGAGAGATATTGTTGGTGC 13057
874 ThrGlyThrProArgLysAlaHisSerLysSerAlaLysAspThrArgAr 890
      ||| ..... |||
13056 TCATCTCTCAACTGAAGAGACATCTCAAACTTCACCCAGCACAGAC 13007
890 gGln...SerGlnProArgThrArgArgSerLysArgThrValProLysP 906
      ||| ..... |||
13006 TCATCTTAAGTCCAGAAAGTGAGACAGATTCATTCAGATACCAATCTC 12957
906 hSerSerGlnLysSerGlnProValArgThrArgArgProArgSerLeu 922
      ||| ..... |||
12956 TGCAGAGCAAAATCTCAAAATCACCCTAAGGAGGTGGTGCAGGTCTTCA 12907
923 ArgLysIleAsnTyrLeuIleAlaArgGluLysAsnGluArgGluLysAr 939
      ||| ..... |||
12906 TCTCCAGTCACTGAGCTGGCATCCAGATCTCCAAATGAACAAGATGAGG 12857
939 gLysLysSerVal.....AspL 945
      ||| ..... |||
12856 TGAATTTCTACGCGAGTCTATGTTCAAATCTGAAATGCTCTCCGACCA 12807
945 ysAspPheGluSerSerSerMetLysIleSerLysValGluGlyThrGlu 961
      ||| ..... |||
12806 GCAGGTTCAGTCTGACTCTCTTCATATCTTACAGTGAAGTCAAGTTC 12757
962 IleValLysProSerProLysArgLysMetGluLysAspValGluLysLe 978
      ||| ..... |||
12756 CTTCTGGGCGCAGAGTAAATGAGACTGCTGAATCAAA...GAGAAAT 12710
978 uGluArgThrProGluLysAspLysIleAlaSerSerThrThrProAla 995
      ||| ..... |||
12709 GGCCTTACCCCTCAGAGAGAT.....GCTACTGCATCACCTCCCTAAGC 12666
995 yslLysIleLysLeuAsn..... 1000
      ||| ..... |||
12665 AGAAAGCAAAATTTAGTCTTCACATACAGGATAGGCTGAGTCTTCA 12616
1001 .....ArgGluThrGlyLysIleGluLysAl 1010
      ||| ..... |||
12615 CTGGTATTTCAAAGACACTTAGAACCCCGCCAGGAAAGAGTGGTGC 12566
1010 aGluAsnAlaSerThrThrLysGluProSerGluLysLeuGluSerThrs 1027
      ||| ..... |||
12565 TGGGTCACTCCAGAAACAAAGAGCAAAATAGTCATTCGCTACCTCA 12516
1027 eSerLysIleLysGlnGluLysValLysGlyLysAlaLysArgLysVal 1043
      ||| ..... |||
12515 GCCAAGATGAAGGTTAAATGAGGTGTAGAGAAAGCTGAAGAACCCGA 12466
1044 AlaGlySerGluLysSerSerThrLeuValAspTyrThrSerThrSe 1060
      ||| ..... |||
12465 GGCCAAATCCTGCTCATTTGCTTCAGAACTTAAAGAAATGTCCACAAG 12416
1060 rSerThrGlyLysSerProValArgLysSerGluGluLysThrAspThrL 1077
      ||| ..... |||
12415 TTACTTTGAATCATCTCTCT.....GAAGTGAAGAAA 12384
1077 ysArgThrValIleLysThrMetGluGluLysHisAsnAspAsnThrAla 1093
      ||| ..... |||
12383 GGCTCTGTGTCTTGTGACTCTGATCAGAGCAGTCAAGCGCTCTTTGG 12334
1094 ProAlaGluAspValIleIleMetIleGlnValProGln.....SerL 1108
      ||| ..... |||
12333 GAAGCA.....GTAAGATCCCTTCATAGGCTTCATC 12302
1108 srPAspLysAspAspPheGluSerGluGluGluAspValLysThrThrG 1125
      ||| ..... |||
12301 TTGGGGTGGCCCATTTTCTCCAGAACATTAAGAACTG...TCTAACT 12255

```


THIS PAGE BLANK (USPTO)

OM of: US-09-811-045A-1 to: Issued_Patents_NA:* out_format: pfs
Date: Sep 12, 2002 4:04 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framed_p2n_model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09811045/rn122242_5997/app_query.fasta.1.1485
-DB=Issued_Patents_NA -QPM=fastap -SUFFIX=ini -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DEPOP=6.000 -DEPEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09811045 -CGN1_1.69 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-811-045A-1
Query length: 1404
Database: Issued_Patents_NA:*
Database sequences: 383533
Search time (sec): 66.770000

score_list:

Sequence	Strid	Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-139-937-11 +			1838.50	1947.00	9.2e-101	1800
/cgn2_6/ptodata/1/ina/PTUS.COMB.seq:PCT-US93-11310-11 +			1838.50	1947.00	9.2e-101	1800
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-931-999-4 +			477.00	481.39	4.0e-19	6755
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-09-036-987A-1 -			366.00	337.68	4.0e-11	80161
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-370-700-1 -			366.00	337.68	4.0e-11	80161
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-428-517-1 -			357.50	333.26	7.1e-11	50937
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-320-559-1 +			343.63	1.9e-11	14255	14255
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-327-392-1 +			343.63	1.9e-11	14255	14255
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-306-618B-55 +			355.00	343.63	1.9e-11	14255
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-545-860D-1 +			355.00	343.63	1.9e-11	14255
/cgn2_6/ptodata/1/ina/PTUS.COMB.seq:PCT-US94-04496-1 +			355.00	343.63	1.9e-11	14255
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-061-376-4 +			345.50	335.34	5.4e-11	11907
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-764-233A-1 +			337.00	311.71	1.1e-09	49377
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-144-085-3 -			329.00	307.14	2.0e-09	33529
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-258-261B-6 +			323.50	302.78	3.5e-09	28958
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-457-342-6 +			323.50	302.78	3.5e-09	28958
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-457-646A-6 +			323.50	302.78	3.5e-09	28958
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-458-076A-6 +			323.50	302.78	3.5e-09	28958
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-764-233A-4 +			323.50	302.78	3.5e-09	28958
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-457-335A-6 +			323.50	302.78	3.5e-09	28958
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-729-214-6 +			323.50	302.78	3.5e-09	28958
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-028-934-6 +			323.50	302.78	3.5e-09	28958
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-769-309A-4 +			322.00	316.31	6.2e-10	6605
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-994-570-4 +			322.00	316.31	6.2e-10	6605
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-09-335-409-1 +			316.50	286.46	2.9e-08	68750
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-568-102-1 -			316.50	286.46	2.9e-08	68750
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-567-969-1 -			316.50	286.46	2.9e-08	68750
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-568-480-1 -			316.50	286.46	2.9e-08	68750
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-568-486-1 -			316.50	286.46	2.9e-08	68750
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-568-472-1 -			316.50	286.46	2.9e-08	68750
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-09-443-501A-2 -			314.00	285.46	3.3e-08	13987
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-804-227C-13 -			314.00	285.46	3.3e-08	13987
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-804-227C-7 -			314.00	288.28	2.3e-08	44377
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-804-198-1 -			314.00	288.28	2.3e-08	44377
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-105-537-30 +			312.50	298.60	6.0e-09	13842
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-105-537-5 +			312.50	288.13	2.3e-08	36778
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-09-130-878-19 +			302.00	291.20	1.6e-08	9551
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-056-200-93 +			302.00	291.20	1.6e-08	9551

/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-728-323A-1 + 301.00 300.44 4.8e-09 3489
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-770-379-20 - 301.00 277.69 8.8e-08 33207
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-757-669A-20 - 301.00 277.69 8.8e-08 3320
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-230-371A-20 - 301.00 277.69 8.8e-08 3220
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-07-642-734C-3 - 300.50 281.92 5.1e-08 20235

seq_name: /cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-139-937-11

seq_documentation_block:
; Sequence 11, Application US/08139937
; Patent No. 5821070
; GENERAL INFORMATION:
; APPLICANT: SHAN, BEI
; APPLICANT: LEE, MEN-HWA
; TITLE OF INVENTION: CELLULAR GENES ENCODING
; TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/139,937
; FILING DATE: 20-OCT-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/979,156
; FILING DATE: 20-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHRYN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 9370
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-139-937-11

alignment_scores:
Quality: 1838.50 Length: 490
Ratio: 4.113 Gaps: 5
Percent Similarity: 91.224 Percent Identity: 75.510

alignment_block:
US-09-811-045A-1 x US-08-139-937-11 ..
Align seq 1/1 to: US-08-139-937-11 from: 1 to: 1800

919 ProArgSerLeu...ArgLysIleAsnTyfLeuIleAlaArgGluLysAs 934
|||||
11 CAAAGAGCCTAATGAGAAACCAACCACTGATATTAAGGAGGAAAA 60
934 ngUARGGluLysArgLys...LysSerValAspLysAspPheGluSer 950
:|||||
61 AGAAAAAGAAAAAGCTGAAGAAAAAGGCGTAGATTAAGATTTTGGCTTC 110
950 rseMetLysIleSerLysValGluGlyTrnGluIleValLysProSerP 967
|||||

```

111 TTCAATGAAATCTCGAAACTAGAAAGTGAATAGTGAACCATCAGC 160
967 rOLyArGLysMetGLuGLyAspValGLuLysLeuGLuArgThrProGLu 983
161 CAAGCGCAAAATGGAACCTGATCTGAAAAATGATGAGACCCCTGAA 210
984 LysAspLysIleAlaSerSerThrThrProAlaLysLysIleLysLeuAs 1000
211 AAGGCAAAAT...TCTTAAAGTGGCCAGCCCAAAAATCAAACTCA 257
1000 nArgGLuThrGLyLysLysIleGLyAsnAlaGLuAsnAlaSerThrL 1017
258 CAGAGAAACTGGGAGAAATTTGGAGATGACAGAAATATATCAAAACCA 307
1017 ySGlProSerGLuLysLeuGLuSerThrSerSerLysIleLysGLu 1033
308 AAGAACCTCTGAAAAATTTGGAGTCAACATCTAGCAAAAGTTAAACAG 357
1034 LysValLysGLyLysAlaLysArgLysValAlaGLySerGLuGLySer 1050
358 AAAGTCAAGAGAAAGTCAGACGAAAGAGTGAAGTGAAGATCCAG 407
1050 rSerThrLeuValAspTyrThrSerThrSerSerThrGLySerPro 1067
408 CTCACCTCTGTGGATTCACCCAGTACGAGCTCACTGGAGGCACTGTG 457
1067 aLArgLysSerGLuGLuLysThrAspThrLysArgThrValIleLysThr 1083
458 TCGGAAATCTGAAAGAAAAACAGATACAAAGCGAACTGTGTTAAAA 507
1084 MetGLuLysTyrAsnAsnAspAsnThrAlaProAlaGLuAspValIle 1100
508 ATGGAAGATATATATATATGACAAATACCGCGCAAGTGAATATCAT 557
1100 eMetIleGLuValProGLuSerLysTyrAspLysAspSerProGLuSer 1117
558 TATGATTTAGGTTCTCTCAATTCGAAATGGATTAAGATGACTTGAAT 607
1117 LuGLuLysAspValLysThrThrGlnProIleGLuSerValGLySer 1133
608 AAGAAGAAAGATTTAAATCCACACAGCCTATATCAAGTGTAGAAAA 657
1134 SerSerIleIleLysAsnValThrThrLysProSerAlaThrAlaLys 1150
658 GCTAGTGTATATAAAAATGTAGTACAAAGCCATCAAAATATATGCA 707
1150 ThrGLuLysGLuSerGLuGlnProGLuLysLeuLysLeuProLys 1167
708 TCTGAGAAAGAAAGTGAAGCCATCCGAGAAATTCAGAAATTCACCA 757
1167 LuAlaSerHisGLuLeuMetGlnHisGLuLeuArgSerSerLysGLy 1183
758 ACGTGAAGCATGAATCATACACATGAGTTAAAAAGTTCAAAAACCT 807
1184 AlaSerSerGLuLysGLyArgAlaLysAspArgGLuHisSerGLySer 1200
808 GCATCTAGTAAAAAGGAAAAACCAAGATCGAGATTTATTCAGTGTGA 857
1200 uLysAspAsnProAspLysArgLysSerGLyAlaGLuProAspLys 1217
858 AAAGGAGAACCTGAAAAAGAGAAAGACAGCATCCAGCCAGAGAAAGA 907
1217 eThrValAspArgLeuSerGLuGlnGLuHisPheLysThrLeuSerG 1233
908 GTAAATTTGACCGTCTGAATGACAGGAAATTTAAAAAGTCTCTCA 957
1234 SerSerLysGLuThrArgThrSerGLuLysHisGLuSerValArgLys 1250
958 TCTTCAGAAAGAGGCTAGAAAGCATGAATCAATGATTCACACGCGCT 1007
1250 rSerAsnLysAspPheThrProGLyArgAspLysLysValAspTyrAs 1267
1008 CTCAAATTAAGACTTCACTCCCAATAGAGACAAAAAACTGACTATGA 1057

```

```

1267 eArgAspTyrSerSerSerLysArgArgAspGLuArgGLuLeuAla 1283
1058 CCAGAGAGATTCAGATTCCAAACCTAGAGATGAAAGATATATTAACA 1107
1284 ArgArgLysAspSerProProArgGLyLysGLuSerLeuSerGLyL 1300
1108 AGACGAAAAAGACTCTCTCTCGAATTAAGATTTCTGCATCTGACAGA 1157
1300 sSerLysLeuArgGLuArgAspLeuProLysLysGLy...AlaGLu 1316
1158 AATTAACCAAGGAGAGAGAGATTTGCTTAAAAAGGAGACGAGATT 1207
1316 eLysLysSerAsnSerSerProProArgAspLysLysProHisAsp 1332
1208 CCAAAAAAGTATTCCTAGTCCCTCAAGAGACAGAAAACTATATCATC 1257
1333 LysAlaProTyrGLuThrLysArgProCysGLuThrLysProValAs 1349
1258 AAAGCCATTATGATATTAACGGCCAAATGAAAGACAAATCTGTAGA 1307
1349 PLYAsnSerGLyLysGLuArgGLuLysHisAlaAlaGLuAlaArgAs 1366
1308 TAAAAATCTTGTAAAGATCGTGAAGAGCATGATATAGACGAAAGACA 1357
1366 LysGLuSerSerGLy...AlaAsnCysHisValTyrLeuThrArgGL 1382
1358 ATAAAGAGTCAGATGCAATTAATCTTATATATCTTATACCCACAG 1407
1382 rLeuProTyrArgArgSerTyrLeuLeuGLyArgTyrPArgAlaPro 1399
1408 ACACAGGTGTAAGAAAGACCAATTAATCTGGCAAAATGCAAGAGT 1457
1399 eSerArgAsnProSer 1404
1458 CAAGCCTAAACCCAGT 1474

seq_name: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: PCT-US93-11310-11
seq_documentation_block:
? Sequence 11, Application PC/TUS9311310
? GENERAL INFORMATION:
? APPLICANT: BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM
? TITLE OF INVENTION: CELLULAR GENES ENCODING
? TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
? NUMBER OF SEQUENCES: 14
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: CAMPBELL AND FLORES
? STREET: 4370 LA JOLLA VILLAGE DRIVE
? CITY: SAN DIEGO
? STATE: CALIFORNIA
? COUNTRY: USA
? ZIP: 92122
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US93/11310
? FILING DATE: 19-NOV-1993
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: CAMPBELL, CATHERYN
? REGISTRATION NUMBER: 31,815
? REFERENCE/DOCKET NUMBER: FP-CJ 9790
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 619-535-8949
? TELEFAX: 619-535-8949
? INFORMATION FOR SEQ ID NO: 11:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1800 base pairs
? TYPE: nucleic acid

```


CORRESPONDENCE ADDRESS:
ADDRESSEE: Hovey, Williams, Timmons & Collins
STREET: 2405 Grand Boulevard, Suite 400
CITY: Kansas City
STATE: Missouri
COUNTRY: U.S.A.
ZIP: 64108

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.3.3
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/931,999
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/710,561
FILING DATE: 19-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26,262
REFERENCE/DOCKET NUMBER: 25043-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 816/474-9050
TELEFAX: 816/474-9057

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 6755 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: UT0007

IS-08-931-999-4

alignment_scores:	
Quality:	477.00
Ratio:	0.958
Percent Similarity:	50.050
	Percent Identity: 21.508

```
alignment_block:
US-09-811-045A-1 x US-08-931-999-4 . .
```

Align seg 1/1 to: US-08-931-999-4 from: 1 to: 6755

[illegible]

3026	CAGGGGACGGGACACGAGAAACACAGAGAAACAGAGCCGACCCGACGAGG	3075
649GluArgLeuLeuProLeuAs	655
3076	CAGCAAAAGGGGGAGAGAAACAAAAAAACAGAAAAACACAGACAC	3125
655	nLLeuArgAsnSerProPheThrArgGLVArgArg	666
3126	CACAGAGACCAACCCAAAGAAACAGAGAAAGAGAGGGGACACAGAA	3175
667GluAspTyr	669
3176	CGAAAAACAAAGCGGCCACACGCCAAACAAAGCCCGCAGGAGAAACAG	3225
670	AlaAlaGlyGlnSerHisArgAsnArgAsnLeuGly	681
3226	GGCAAAAGGGGCAAGAGCGAGCGACAGCAAAAGCAAAAGCAAGCCGC	3275
682GlyAsnTyrProGluTysLeuSerThrArgAspSer	693
3276	MACCCAAAAGAGAGAAAGCACCCACGAAAGCGCAGCAGACAGAAACAG	3325
694	HisAsnAlaTys..AspAsnProTysSerTysGlyGlnSerGluAs	709
3326	CACGAGAAAAAACACCAACAAACAGACACGACAGCAAGACAGAAAAAG	3375
709	nValProGlyAspGlyTysGlyAsnLysHisLysLysHisArgLysArg	726
3376	GAGGACAGGAGAAAGAAAGGGCGCACAAAGAAAGAAAGAAAGAA	3425
726	ArgAsnGlnGluLys.....GlyGlnGlnSerGlnSerPheLeuAsn	739
3426	ACAAAGCCGAAAAAGAGAGGGAGAGAAAGAAAGAAAGAAACAA	3475
740	ProGlnLeuLeuGlnThrSerArgTysCysArgGlySerSerGlyLeAs	756
3476	ACCGAGCAGACAGAAAAAAAGGGGAAAAAAACAAACAAAAAAACA	3525
756	pcLutHrTyrThrAspThrLeuPheValLeuProSerArgAspAlaT	773
3526	GAAAGACAGACAGAA.....	3541
773	hrProValArgAspLupPonethAspAlaGlnSerIleThrPheLysSer	789
3542	..GAAAAAAAMAAACAGACAGCAGCAACGAAAGAAACAAAGCGAAAC	3589
790	ValSerAspLysAspLysArgGlnLysAspLysPro.....	801
3590	GAAAAAACGAGAACGAGAGAAAAAAACCCACAGCACACCGCAG	3639
802LysValLysSerAspLysThrLysArgLysS	812
3640	CGACCAAAAAAAAGCGCGCAACCAACAGAGAAAAAAACCAAAAGAGGA	3689
812	erAspGlySerAlaThrAlaLysLysAspAsnValLeuLysProSerLys	828
3690	AAGAAAAAAACGAGCAGCAAAAA.....GCCAGCAAAAA	3727
829	GlyProGlnGlnLysValAspLysPArgGlnLysSerProArgSerGln	845
3728	GGAGAGAACAGAAAAAACCAAAAAAGGGGAAAAAGCAAAAAACGGCGGA	3777
845	uproProLeuLysLysAlaLysGlnGlnAlaThrLysLysLeuAspSerValL	862
3778	CAGACCCACAAAAAAAGGAGCAGAGAAAGCAAAAGAAAGAAACCAAC	3827
862	ysProSerSerSer.....SerGlnLysAspGlnLysValThrGly	875
3828	AAAAAGAGAGCGCAAAAGAAAGAAACAGAGAGCGCAAAAAACCGCGC	3877
876	ThrProArg.....LysAlaHisLe	882
878	ACAAAGAAAGAGAGAGAGAGGCAACGAAACGAAACCAAAAGAGAC	3927


```

seq_documentation_block:
: Sequence 1, Application US/09036987A
: Patent No. 6143526
: GENERAL INFORMATION:
: APPLICANT: Baltz, Richard H.
: APPLICANT: Crawford, Mary C.
: APPLICANT: Broughdon, Kathryn P.
: APPLICANT: Madduri, Krishnamurthy
: APPLICANT: Merlo, Donald J.
: APPLICANT: Treadway, Patil J.
: APPLICANT: Turner, Jan R.
: TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
: TITLE OF INVENTION: Production
: NUMBER OF SEQUENCES: 39
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dow Agrosciences LLC Patent Department
: STREET: 9330 Zionsville Road
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: USA
: ZIP: 46268
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/036,987A
: FILING DATE: 09-MAR-1998
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Stuart, Donald R
: REGISTRATION NUMBER: 28,479
: REFERENCE/DOCKET NUMBER: 50, 608
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (317)337-4816
: TELEFAX: (317)337-4847
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 80161 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-09-036-987A-1

alignment_scores:
Quality: 366.00 Length: 1708
Ratio: 0.499 Gaps: 83
Percent Similarity: 42.974 Percent Identity: 19.965

alignment_block:
US-09-811-045A-1 x US-09-036-987A-1/rev ..
Align seg 1/1 to reverse of: US-09-036-987A-1 from: 1 to: 80161

23 ProthriIleasPAlaGlu.....AlaTyraIaIleGlyLysGly 37
|||||
67729 CCGACCATTCGACCGAACCACACCGCGTGTGGCCGCCGCCACCGCGCG 67680
37 sProthPhe.....L 41
|||||
67679 ACCACAGTTTCGTTCCGTACGGTCTTTACGACGACATCGGATCGCAT 67630
41 euProGluGluProSerSerSerSerGluGluAspAspProIleProIa 57
|||||
67629 GCGCGATGACGCGACGATGACATCGGGAATCAGCGCGCAC 67584
58 GluLeuLeuGlyLeuIleCysLysAspIleMetThrAspAlaValIle 74
|||||

```

```

67583 .....GTCCATCAGGGCCGACGAGTTCGAAGATCGTGCGAC 67549
74 eProCysCysGlyAsnSerSerCysAspGluCysIleArgThr..... 88
|||||
67548 CCCTTGCTGCGGAGACCCGAGCCCGCTGGCGAAGCGGAGCTCTC 67499
89 .....Thr..LeuLeuIleSerAspLysHisThrCys..... 98
|||||
67498 GAACCTGACGACCCCATATCTCCGGCGTAGCCATCAGACCTCGGGTTCG 67449
99 .....ProThrCys 101
|||||
67448 AGCTACCCGTCACGTCAGACGACGCGTAGTGGCCAGCGCGTAGTC 67399
102 HisGlnAsnAspValSerProAspAlaLeuIleAlaAsnLysPheLeuAr 118
|||||
67398 CACGCT.....GCGTCCGATTCGGGAACTCTGTGA 67367
118 gGlnAlaValAsnAsnPhelysAsnGluThrGlyTyThrLysArgLeuA 135
|||||
67366 ACATCGGCTCCATCCGATACGA.....ATGAACCGCATTC 67332
135 IglysGlnLeuProPheLeuPheLeuValProProArgProLeu 151
|||||
67331 GAACCCCGAACCG.....CCGCCACCGAATCCCTTG 67300
152 SerGlnArgAsnLeuGlnProArgSerArgSerProIleLeuArgGlnI 168
|||||
67299 .....CCATTCACGCGCGCG 67283
168 nAspProValValPheArgTyThrValSerProThrCysSerAspThrL 185
|||||
67282 CGATGTGCTC.....AGCACCTCC 67263
185 ySThrAlaGly..SerCysSerSerGlyThrLeu...SerArgLeuPr 200
|||||
67262 CGATCGCGGAGAGCACACGACCTAGAGCGGTGACCGCGCATAC 67213
200 oAlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAlaProP 217
|||||
67212 CCGACCATGACCGCATCGCGCCACACGCGCGAGCTGCTCTTACACG 67163
217 roValSerGlyAsnProSerSerSerAlaProAlaProValProAspIleThr 233
|||||
67162 CAGCCACCGCAC.....ATGCACCG.....CCAGAAGGCAAC 67128
234 AlaThrValSerIleSerValHisSerGluLysSerAspGlyProPheAr 250
|||||
67127 GCCTGCTATC..... 67119
250 gasPserAspAsnLysLeuLeuProAla.....AlaAlaLeuThrS 264
|||||
67118 .....AACGGGCGACGACCGCCACCATTCGCGAGCGCATCCGGCA 67079
264 eArgLHisSerLysGlyAlaIleSerSerIleAlaIleThrAlaLeuMetGlu 280
|||||
67078 ACGACACACCTCCCGCGCGACCGCGCGCGGACGTAATCGCGCACGAATGG 67029
281 GlnLysGlyValProGlyThrSerProThrAsnSerIlePheValGlyI 297
|||||
67028 CCCACACCCACCGCGCGCGCGCACCGA..... 66996
297 nSerLeuLeuHisGlyGlnLeuIleProThrThrGlyProValArgIleA 314
|||||
66995 .....CCACGACGCTCA 66983
314 snAlaAlaArgProGlyGlyAlaArgProGlyTyThrGlnHisSerAsnLys 330
|||||
66982 GCAACCAACCTGCAACCGCAACCAACCGGACTGGGCCACAGCGCTCGCA 66933
331 LeuGlyTyThrLeuValSerProProGlnGlnIleArgArgGlyGluArgSe 347
|||||
66932 TCCACAAGAGCCCGTGGAAACCAACACACATCCGCGACCCCGACGTC 66883

```



```

347  rCysTyrArSerIleAsnArgGlyArgHisHisSerGluArgSer  Gln  363
      ||||| ||||| :::: ||||| ||||| :::: |||||
66882  CTGCCCCAGATCGCGCTCCAGTCGGCAGACATCTTGTCATAAGCGGCTCG  66833
364  ArgThrGlnSerProSerIleuProAlaThrProCysPheValProValPr  380
      ||||| ||||| ||||| ||||| :::: |||||
66832  CGAACA.....CCGAAAAACCGAGTAGAGCCCCGGCCCC  66798
360  oProProProLeuTyrProProProHisThrIleuProIleuPro...  395
      ||||| ||||| :::: |||||
66797  ATCCCAACCACTGTCCGCCCTGACCGCGGACACGAACCAACCCCGCC  66748
396  ....ProGlyValProProGlnPheSerProGlnPheProSerSer  410
      ||||| ||||| ||||| |||||
66747  AGAGACCGGGATCCCGACACA.....CCCCAGAGCCTCGCGAC  66707
411  GlnProPro.....ThrAlaGlyTyrSerValProProPr  422
      ||||| ||||| ||||| |||||
66706  CGGAGCCGAGCGCTTCCAACCGGACACAGATTTCGGCGTCCGCTGCC  66657
422  oGlyPheProProAlaProAlaAsnIleSerThrAlaCysPheSerProG  439
      ||||| ||||| ||||| |||||
66656  AGCACACCGCGCCGCTCTCCACAG.....CGGCGCG  66625
439  ly...ValProThr.....Ala  443
      ||||| |||||
66654  GGAGAGACCAAGCAAAACCGCGCATCCAGCGTGGAAATATCGCCAGCT  66575
444  HisSerAsn.....ThrMetProThrThrGln.....  452
      ||||| ||||| ||||| |||||
66574  CACCCAAATAGACAACAATGCGCTGAGCGCTGAGCGCTGCGCGT  66525
453  ...AlaProLeuLeuSerArgGluGluPheTyrArgGluGlnAsnAspL  468
      ||||| ||||| |||||
66524  GTTCTCCCCGACACGATCAACGGGACCGCTGGAAGTCCACGCTGATTC  66475
468  ysgLyArgGluSerLysPheProTyrSerGlySerSerTyrSerArg...  483
      ||||| ||||| ||||| |||||
66474  CGAATCGGGACGCTGTGTCCGACGTCTGTCGTGTTCGAGGCGTT  66425
484  .....SerSerTy  486
      ||||| |||||
66424  GTTCGAGGATGACGTGCTGCTGTCGCGCTGATCCCGACGACGACGACG  66375
486  rThrAspSerSerGlnGlyLeuAlaGlnHisIleHisAlaLeuThrLeuS  503
      ||||| ||||| ||||| |||||
66374  CCCACCCCGCGGAGATGGCTGTTCGCG.....GCCAGGGGGTGT  66334
503  ePrProSerAlaAlaHisThrLeuAspLeuLeuHisAsp.....  515
      ||||| ||||| |||||
66333  CTCCCTGAGGAGCTGGA.....CATCCCGCGCGACCAAT  66299
516  .....HisProHisProProGlnGluAlaGluAlaIar  526
      ||||| ||||| ||||| |||||
66298  CCACCTCCGACCTAGGTTTCATCCACATGCACTGCGGGCAGCTGCGCG  66249
526  gSerAlaMetIleVal.....HisMetProAspL  536
      ||||| ||||| ||||| |||||
66248  TCCCGCATGCGCATCACCATTGTGATCACAACGAGGACACCGCGCGCG  66199
536  euMetAspIleAlaHisAlaArgSerArgSerPro.....  547
      ||||| ||||| |||||
66198  CTGCGTATGACCAATATTCGACTTCACCAACCCCAACCAACGCGCAT  66149
548  .....ProTyrArgArgTyrArgSer.....  554
      ||||| ||||| ||||| |||||
66148  CGCTATCCCGCCCTGCGCGTAAAGTGGCAATCAGAGACCTGCGCGCTCA  66099
555  .....ArgSerArgSerProProGluPheArgGlyG  565
      ||||| ||||| ||||| |||||
66098  GGATCACCAGCAAGCGTGGTTCAGTCCCATGCGCTCCAGCGGCGTCCAC  66049

```

```

565  lSerProThrLysArgAsnValPro.....  573
      ||||| ||||| ||||| |||||
66048  AGACACCGCAAAACCAAGCAGCTGCCACAGCTGCGGATACACCGCTGCT  65999
574  ...ArgGluGluLysGluArgGluTyrPheAsnArg.....T  585
      ||||| ||||| ||||| |||||
65998  GCGAGGCGCCCATTCGGCGCGCTCAAAACCATTCGACGACCAACATCCTG  65949
585  yTrArgGluValProProProTyrAspIleLysAlaTyrTyrGlyArgSer  601
      ||||| ||||| ||||| |||||
65948  ACCGCGCTAAC.....CCGACAGACCGGCAAGACCGCGCGCGTGG  65905
602  ValAspPheArgAspProPheGluLysGluArgTyrArgGluTyrPolar  618
      ||||| ||||| ||||| |||||
65904  CCGGGCATCCGAGCGCGCTCCACACACCAAGCCCGCGC.....  65863
618  gLysTyrArgGluTyrPyrGluLysTyrTyrLysGlyTyrAlaValGlyA  635
      ||||| ||||| ||||| |||||
65863  .....  65863
635  lAglnProArgProSerAlaAsnArgGluAspPheSerPro.....  648
      ||||| ||||| |||||
65862  .....TTCCTCCCAACCGGTGCC  65845
649  .....GluArgLeuLeuProLeuAsnIleArgAsnSerPr  660
      ||||| ||||| ||||| |||||
65844  ATCCGACGCCCGCGCAACGACTTGCACCGCCCATCCGCCCAAGCCCC  65795
660  oPheThrArgGlyArgArgGluAspTyrAlaAlaGlyGlnSerHisArgA  677
      ||||| ||||| ||||| |||||
65794  GTTGACGCCAATAATCCAGCAACATCCCGCGGTGGCAT...CACCGTC  65748
677  snArgAsnLeuGlyLysnTyrProGluLysLeuSerThrArgAspSer  693
      ||||| ||||| |||||
65747  AC.....ACCACCGCAAGAC  65731
694  HisAsnAlaLysAspAsnProLysSerLysGluLysGluSerGluAsnA  710
      ||||| ||||| ||||| |||||
65730  CAGATCACACTACGACGACGAGCTGTGACCCGCGCAGGTGAG.....  65686
710  lProGlyAspGlyLysGlyAsnLysHisLysLysHisArgLysArgArg  726
      ||||| ||||| ||||| |||||
65685  .....CGCCACCAAGCAGCAGCAGACACGCGCTA  65658
727  ....AsnGluGluLysGlyGluGluSerGluSerPheLeuAsnProGlu  741
      ||||| ||||| ||||| |||||
65657  TCCACCGTACCGCAGAGACCTCGAAACCAACGAAATAGCAACCCGAC  65608
742  .....LeuLeuGluThrPhe  746
      ||||| ||||| |||||
65607  CGAAAAACAGCGCTCCGCGCTGCCATTACCTAGATTAACCTCGAAACCT  65558
746  rArgLysCysArgGlySerSerGlyIleAspArgLysThrLysThrAspTr  763
      ||||| ||||| ||||| |||||
65557  CCGGCGCCCTCT.....GATGAACCGCGCGCGCTGATGTCG  65523
763  euPheValLeuProSerArgAspAspAlaThrProValArgAspGluPr  779
      ||||| ||||| ||||| |||||
65522  TGGTGCATGAGCCCGCGAAC.....ACGCGGTCCGCTCCCGCG  65482
779  oMetAspAlaGluSerIleThrPheLys.....  788
      ||||| ||||| |||||
65481  CAACGAAACGGATCAATCCCGCGCTCCACGCGTTTCCCAAGCAGATT  65432
789  .....SerValSerAspLysAspLysArgGlu  797
      ||||| ||||| ||||| |||||
65431  CAGACGCAACGCGCTGCTGCGATCCATCGCAACGCTCAGCGGGGAA  65382
798  ...LysAspLysProLysValLysSerAspLysThrLysArgLys.....  811
      ||||| ||||| ||||| |||||
65381  ATCCGAAAGAAACCGGATCAGAACTCTGCCGATCTCGAGGAAAGCACCC  65332
811  .....  811

```

```
65331 CTCACACATAGAGTCCCAAGACGATCGGATCCGATCGAACAACC 65282
812 .....SerAspGlySerAlaThrAlaLysLys 820
65281 CCGCCTGATCCACGACGATGTCGGGAATCCGCAACGCGATCCAAAG 65232
821 AspaSnValLeuLysProSer..... 827
65231 CCCCCGGCCACCAAGCCAAAGCTCTCCGGCGAGACACCCCAACCCGG 65182
828 ...Lys.GlyProGlnGluValAspGlyAspArgGluLysSerPro 842
65181 ATAAAGACAGCCCATCCCAAGATCACAAACGGATCGTCAGACAGTACCG 65132
843 ArgSerGluProProLeuLys..... 849
65131 CCGCCACAGCCGCGTCCGGAGTCCGAGCTCCACGCCCAATCATAGT 65082
850 ...LysAlaLysGluGluAlaThrLysIleAspSerValLysProSerS 865
65081 TTGTCAGACGAAGAACCAACGACG....CCGCGGTGTCCGATATGTC 65038
865 erSerSerGlnLysAspGluLysValThrGlyThrProArgLysAlaHis 881
65037 GAACACGCGCATAGCCGCGCAACGACGCCCAACCCCGCAAGCCGCT 64988
882 SerLysSerAlaLysAspThrArgArgLysSer..... 892
64987 TCCGAACATCCACACCGACACAGATATAAACCAATCCCGGAACGA 64938
893 .....GlnProArgThrArgArgSerLysArgThrValProLys. 905
64937 CGATCGGCCCGACCTGACCGCAACCCGAAATGCCACACCGCGGAAC 64888
906 .....Thr 906
64887 ATGCTCGGCAACAACCAACAGCTGGTTCTGATCCAGGTCCGCGCA 64838
907 SerSerGlnLysSerGlnProValArg.....ThrArgArgPr 919
64837 GCCTCCGACGCGCAACAACGACCGTTGACGTCCGCGACTCGTTGACC 64788
919 OAArgSerLeuArgLysIleAsn.TyrLeuIleAlaArgGluLysAsnGlu 935
64787 GCGCTCCGCGCCAGCGCGGTGCGACAGCGTCCGCAACAACGACGCGAC 64738
936 ArgGluLysArgLys.....LysSerValAspLysAspPheGluSe 949
64737 ACCGCGCTCGCGCATCAAGCGCGAGGGCGTTCAGTCCAGCTGGACCG 64688
949 rSerSerMetLysIleSerLysValGluGlyThrGluIleValLysProS 966
64687 GAACCAAGAGGATTCGCCCGGAACGCAACGCGCATCTCGAACAACGCGC 64638
966 erProLysArgLysMetGlnGluLysPValGluLysLeuGluArgThrPro 982
64637 AACCCGTCCGCGCGGACATCGCGGATTTGTT..... 64606
983 GluLysAspLysIleAlaSerSerThrThrProAlaLysLysIleLysLe 999
64605 .....GCGCGCAATGCGGCTACGCGCCGCGCATCCGACGCG 64571
999 uAsnArgGluThrGlyLysLysIleGlyAsnAlaGluAsnAlaSerThrT 1016
64570 CTCGCGTCATCCGCTGCGGTCTCCCAAGCCCAAGCCAGCGAGACACC 64521
1016 hrLysGluProSerGluLysLeuGluLysThrSerSerLysIleLysGln 1032
64520 GCAGCGACGACC....CGCGTCCGCGGTGTCGAGCCAGCGCGTCCAG 64477
1033 GluLysValLysGlyLysAla.....LysArgLysValAla 1044
|||||:|||||:.....:|
```

```
64476 GAACGCGTTCCGCCCGCGGTAGTTTCCCTGACCCGCAACCCCGCCACGC 64427
1044 agLysSerGluLysSer..SerSerThrLeuValAspTyrThrSerThrSe 1060
64426 CGGAGAACGAGAGAGAGACAGCAAGAAAGACAGAGATCCAGATCAAGAGTC 64377
1060 rSerThrGlyLysSerProValArgLysSerGluGluLysThrAspThrL 1077
64376 AACTGTCCAGAGTGCACACGACCGCGGCTTCGGGGCCGAACACCTTTCG 64327
1077 ysArgThrValIleLysThrMetGluIuTyrAsnAsnAspAsnThrAla 1093
64326 CACGTGCTCTGTGTCTC.....AGCGATCCAGCGACC 64296
1094 ProAlaGluAspValIleIleMetIleGlnValProGlnSerLysTyrPas 1110
64295 CCGTGGCGAATGCCCCCGCCGATGCACCACTCCGCGACGCGGGAATCTC 64246
1110 pLysAspAspPheGluSerGluGluLysAsp.....ValLysThrT 1124
64245 TGCCGAGATGTCGTGTCAGACACCGGTCCCAATGGGCCCCGATCCCGACAT 64196
1124 hrGlnProIleGlnSerValGlyLysPro.....SerSerIleLe 1137
64195 CGCAGCTCCGACCTCGACCGCAGCTCCAGGCCCAACCACTCATCAC 64146
1138 LysAsnValThr.....ThrLysProSerAlaThrAl 1148
64145 AGTCGCGGACCCCGGGGGCGCTCCAGCGCGGTGCGCCCGGAGACACAG 64096
1148 aLysTyrThrGluLysGluSerGluGln..... 1157
64095 GCGCGCGACACCGCGCTCCGCGCAACAATGCGGGCAACAACCGGCCCA 64046
1158 .....ProGluLysLeuGlnLysLeuProLysGluAlaSerHis 1170
64045 GCAGCGCCGTAACCGCGGAATCAACACCTGCCA..... 64011
1171 GluLeuMetGlnHisGluLeuArgSerSerLysGlySerAlaSerSerG 1187
64010 .....TCCGCAACGGGCGACTGGGATCCGGCT 63985
1187 uLysGlyArgAlaLysAspArgGlnHisSerGlySerGluLysAspAsn 1204
63984 GTCCTCGCGCGCAACACATCCGCGCAGGCGAGCGAC....CGGACTTCAC 63938
1204 roAspLysArgLysSerGlyAlaGlnProAspLys..... 1215
63937 CCGCGCGCAACGCGAGCTCGCGGTCTCCGGAAGCAGACGCGCGCGCAC 63888
1216 .....GluSerThrValAspArgLeuSerGluGlnGlnHisPheL 1229
63887 GCAGCGCATGATTCGGCGGTTCGTCACATCGAACCAAGCATTCGCTC 63838
1229 sThrLeuSerGlnSerSerLysGlu.ThrArgThrSerGluLysHisGlu 1245
63837 CGGGTTCCTCCAGTCCGCGCAACCAACCAACCCACGACGACGATGCA 63788
1246 SerValArgGlySerSerAsnLysAspPheThr..... 1256
63787 CCAGCTGGGAACACCTCGCGGGAGCTGTCCGAGCGCGCCGACCAACGTC 63738
1257 .....ProGlyArgAspLysLysValAspTyrAspSerArgAspTyrS 1271
63737 ACAACCAACAGGCGAGATCCGGCAATCCGCGCGCCCGCAACCAATCTCG 63688
1271 erSerSer.....LysArgArg 1276
63687 GATCAGGTCACAGATGCGTTGTGTCCGGGACAGTGTCCGGAACCAAGCT 63638
1277 AspGluArgGlyGluLeuAlaArgArgLysAspSerProProArg..... 1291
63637 CCAAGTCTCTGCGAGACGACCGCATG....GCACCACACGACGACTCC 63594
```

```

1292 ..... GlyLysGluSerLeuSerGly..... 1298
1293 ..... |||||
13535 GGGGGGGGTGACCCACCCGGGTGAGTGGGGTGCAGTCCGTCGCAAG 63544
1299 ..... GlnLysSerLysLeuArgLysGluArgAspLeuProLys 1311
1299 ..... |||||
135343 TGCCGAGCCGAGAGCGGTGGGGGAGTTCGTCATTCGCATTCACGTGCCCAAG 63494
1312 LysGluAlaGluSerLysLys.....SerAsnSerProProAr 1325
1312 ..... |||||
13493 AGACCGCGTTGGCCACGCGCGGCGAGATTGCGTTCAGTCGACCCGGAAAC 63444
1325 GAspLysLysProHisAspHisLysAlaProTyrGluThrLysArgProC 1342
1325 ..... |||||
13443 AGGACATCGTGTTCATCGCAC..GGGCGCTGCCAACTGCTTCCGA 63397
1342 ysgLgluThrLysProValAspLys.....AsnSer 1352
1342 ..... |||||
13396 AATCGGGCGACGACACACGATCGATCGATGAGACAGCGGCTTGGCCAGTCG 63347
1353 GlyLys GlnUArgGlyLysHisAlaAlaGlnLysArgAsnGlyLysGluS 1369
13346 TGTCGGTGTGACACCGACACCGCATTCGGCGCCGCGGAACGAGGTGC 63297
1369 erSerGlyAlaAsnCysHisValTyrLeuThrArgGlnThrLeuProTrp 1385
13296 ACCCGAGGGAAGAAAGTCGCGCGCTTCGTTGAGGACACACCTTGCACAGC 63247
1386 ArgArg 1387
1386 ..... |||||
63246 GAAAGG 63241

```

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-370-700-1

```

seq.documentacion.block:
? Sequence 1, Application US/09370700
? Patent No. 6274350
? GENERAL INFORMATION:
? APPLICANT: Baltz, Richard H
? APPLICANT: Broughton, Mary C
? APPLICANT: Crawford, Kathryn P
? APPLICANT: Madduri, Krishnamurthy
? APPLICANT: Treadway, Patli J
? APPLICANT: Turner, Jan R
? APPLICANT: Waldron, Clive
? TITLE OF INVENTION: Blosynthetic Genes For Spinosyn Insecticide
? FILE REFERENCE: 50489 DIvl
? CURRENT APPLICATION NUMBER: US/09/370,700
? CURRENT FILING DATE: 1999-08-09
? EARLIER APPLICATION NUMBER: US 09/36987
? EARLIER FILING DATE: 1998-03-09
? NUMBER OF SEQ ID NOS: 39
? SOFTWARE: Patentln Ver. 2.0
? SEQ ID NO 1
? LENGTH: 80161
? TYPE: DNA
? ORGANISM: Saccharopolyspora spinosa
? US-09-370-700-1

```

	alignment_scores:	
Quality:	366.00	Length: 1708
Ratio:	0.499	Gaps: 83
Percent Similarity:	42.974	Percent Identity: 19.965

alignment_block:
US-09-811-045A-1 x US-09-370-700-1/rev

Align seg 1/1 to reverse of: US-09-370-700-1 from: 1 to: 80161

23 ProThrIleaspalalu.....AlaTyraIaIeGlyLysLysGlu 37
 ||||| ||| ||| :: ::::

67729	CCGACCAATTCAGACGGGAACACCAGCGGTGTGGCCCCGGGCACCGCGGGC	67680
37	sProProbe.....L 41	
67679	ACCAAGATTCCGTTTCCGTACGGCTTTACGACGATCGAANTCGCAT	67630
41	eUProGLUGLProSerSerSerSerSerGLUGLInAspAProlleProla 57	
67629	GCGCGATCAGGCCAACGAAATGCATCGGAATCGAGGCGCGCAC... 67584	
58	GIuleuLeuCysLeuIleCysLysAspIleMetThrAspAlaValI 74	
67583GTGCATCAGGGCCGCGACTTGAAATCGTGGCGAC 67549	
74	eProcYscysGIyasnSerSerCysAspILucYsIIeargthr..... 88	
67548	CCCTTGTCTCCGGAGACACCGGACCGCGCGAAGGAGGAGGCGCTTC 67499	
89Thr..LeuLeuGLSerAspLysHisThrCys..... 98	
67498	GAACCTGACGACCCAAATCTCGCGGTAGCCATCACACTCGCGGTGCG 67449	
99ProthrCys 101	
67448	AGCTACCCGTCACAGCTGACAGCACCGGTAGTCCAGCGCGGTATGTC 67399	
102	HISGLInAsnAspValSerProAspAlaLeuIleAlaAslySphLeuar 118	
67398	CACGCT.....GCGTCCGATTTCCGGGAACTCTCGA 67367	
118	gGLInAlaValasnaSnPheLysAsnGLuthrGLyThrILysArgLeua 135	
67366	ACATCGGGGTCATCGCATACGA.....ATGAAMCCCATGTC 67332	
135	rGLyGLInleuProProPheLeuPheLeuValProProArgrProleu 151	
67331	GAACCCCGCACCG.....CCGCGCACCGAATCCCTTG 67300	
152	SerGLInArgsnLeuGLInProArgSerArgSerProIleLeuArgGLnI 168	
67299CCCATCCACGCGCGCG 67283	
168	nAsProValValPheArgTYrThrValSerProThrcysSerAspThrL 185	
67282	CGATCGCTC.....AAGCACCTTC 67263	
185	ySThrAlaGLy..SerCysSerAspSerGLyThrLeu..SerArgLeuPr 200	
67262	CGATCGCGGAGAAGACCAACCGATCTAGAGAGGTGTACCGCGCATATCC 67213	
200	oAlaProserILeserSerLeuThrSerAsnGLInSerSerLeualProp 217	
67212	CGCACCATCAACCCGATCGGCGCAACAAGCGCGCAGCTCTCTCAACAG 67153	
217	rovalserGLynsProSerSerSerAlaProAlaProValProaspIlethr 233	
67162	CAGCCACCGCCAAC.....ATGGCACCG.....CCAGAAGGCAAC 67128	
234	AlaThrValSerILeserValHisserGLyLSerAspGLyProphear 250	
67127	GCCTGCATC..... 67119	
250	gaspsersAspAslySLeuleuProLa.....AlaAlaLeuThrs 264	
67118AACCGGGCACGACCGCGCCACCTCCGAGCGCATCCGGCA 67079	
264	eergLHIsSerLySLyalaSerSerILealaleThrAlaLeuMetGLu 280	
67078	ACGACAACACTCCCGCGCAACCGCGCGGAGACTCGCGCACCGAATGG 67029	
281	GLyLySGlyValProGLyThrSerProTrpAsnSerILlePheValcylgl 297	
67028	CCCACACACACACCGCGCGGACACCCACGAA..... 66966	

281 **GL**uysgIyVal**Pro**GIy**Thr**Ser**Pro**TrpAsnSerIle**phe**Val**Gly** 297
 :::::||||| ::||| :::
 57028 CCCAGACCACACCCGGCCGACACCCACGAA..... 66996

297 nserleuLeuHisgluInleuIleProThrThrGlyProValArgIleA 314
66995
314 snAlaIaAlaProGlyGlyIleArgProGlyIleHisSerAnlys 330
66982 CGAACCACCACTGCACACCGAACAACCCGAGCTGGCCACAGGCTCCGA 66933
331 LeuGlyTyrIleValSerProProGlnGlnIleArgArgGlyIleArgS 347
66932 TCACAACAAGACCCGTCGGAACAACAACAACATCCCGACCCCGACGTC 66883
347 TCysTyrArgSerIleasnArgGlyArgHisHisSerGlyArgSer, Gln 363
66882 CTGCCCGCCAGATGGCGGTTCAGTCCGGCCACATGCTTCGTCACAACGGGTAG 66833
364 ArgThrGlnSerProSerLeuProAlaThrProCysPheValProValPr 380
66832 CGAACCA.....CCGGAAACACGAGTAGAGCCCGCGGCC 66798
380 oProProleuTyrProProProProHisThrIleuProleuPro.... 395
66797 ATCCCCACCACTGTCGCCCTGACCGCGGACACGAAACCAACCCCGCC 66748
396ProGlyValProProProGlnPheSerProGlnPheProSerSer 410
66747 AGACACGGGATCCCGACCA.....CCCGAAGCCTCGCGAC 66707
411 GlnProPro.....ThrAlaGlyTyrSerValProProPr 422
66706 CGGAACCCGCTTCAACCCGGAACAACGTTCCGGTCCGCTCC 66657
422 oGlyPheProProAlaProAlaAsnIleSerThrAlaCysPheSerProG 439
66656 AGCACACCGCCGCTCTCCAACG.....CGGCCG 66625
439 Ly.....ValProThr.....Ala 443
66624 GGAGAGACCAACGAACGCGCATCCAGCGTGGAATATCGCACGCT 66575
444 HisSerAsn.....TherProThrThrGln..... 452
66574 CACCCAAATGAGACAAATCGCTCGCTGAGCGCTGAGCGCTCCGCT 66525
453AlaProleuLeuSerArgGlnGluPheTyrArgGlnAsnAspRL 468
66524 GTTCTCCCGACAGATCAACGGACCGCTGAAAGTCCACGGTCGATTTC 66475
468 ysgIyArgGluSerLysPheProTyrSerGlySerTyrSerArg... 483
66474 CGAATCGGGACCGCTGTGTCGCGAGTCTCGTGTGTTTCGAGGCTT 66425
484SerSerTy 486
66424 GTTCGAGAGTACGCTGCTGTGCTGCTGATCCCGAAGCAGCAGCAG 66375
486 rThiAspSerSerGlnGlyLeuAlaGlnHisIleHisAlaLeuThrIleus 503
66374 CCCAACCGCGGAGGTGCTGTGCGG.....GCCAGGGGGTGT 66334
503 erProSerAlaAlaHisThrIleuAspLeuIleHisAsp..... 515
66333 CTCGGTAGAGAGCTGA.....CATCCCGCGCGCAGCAAT 66299
516HisProHisProGlnGluAlaGluAlaAr 526
66298 CCACTTCGAGTAGGTGATCATGCATGCAAGTCGCGGCAAGTCGCG 66249
526 gSerAlaMetIleVal.....HisMetProAspRL 536
66248 TGGCGCATCGCCATTCACCATCTTGATCAGACAGGAGACCCGCGCGC 66199

536 eumElAspIleAlaHisAlaArgSerArgSerPro..... 547
66198 CTGGGTATGACCAATATGACTTACACCGACCCCAACCAAGCGCGAT 66149
548ProTyrArgArgTyrArgSer..... 554
66148 CGGTATCCCGCCCTGCGCGGTAGTGGCAATCAGAGCTCGGCTCAATC 66099
555ArgSerArgSerProProGlnPheArgGly 565
66098 GGATCACCACCGCTGTTCCAGTCCATGCGCTCCAGCGGTCCACATC 66049
565 InSerProThrLysArgAsnValPro..... 573
66048 AGACACCGCAACACGACATCGCCAAACGCTGCGTATCAGCCGTCGT 65999
574 ...ArgGluGlu, LysGluArgGluTyrPheAsnArg.....T 585
65998 GCGAGGGCCCTTCGCGCGCTCAACCATTCGACGACACCATCTGATTTC 65949
585 yTArgGluValProProProTyrAspIleLysAlaTyrTyrGlyArgSer 601
65948 ACCCGCTAC.....CCGCACAGCGCCGAGAACCGCGTCCGCTTCG 65905
602 ValAspPheArgAspProPheGlnLysGluArgTyrArgGluTyrGluAr 618
65904 CCGGACATCCGACAGCGCTCCAAACAACACCAACCCCGCCG..... 65863
618 gLysTyrArgGluTyrPyrGluLysTyrTyrLysGlyTyrAlaValGlyA 635
65863 65863
635 InGlnProArgProSerAlaAsnArgIleAspPheSerPro..... 648
65862TTCCTCCCAACCGGTGCC 65845
649GluArgLeuLeuProLeuAsnIleArgAsnSerPr 660
65844 ATCCGACCGCCCGCAACGACTTCGACCGCCATCCCGCCCAAGCCG 65795
660 oPheThrArgGlyArgArgGluAspTyrAlaGlyGlnSerHisArgA 677
65794 GTTACCGGAAACTCCACGAACATCCCGCGGTGGCAT...CACGCTC 65748
677 snArgAsnLeuGlyGlyAsnTyrProGluLysLeuSerThrArgAspSer 693
65747 AC.....ACCACCCGCAAGAC 65731
694 HisAsnAlaLysAspAsnProLysSerLysGluLysGluSerGluAsnVa 710
65730 CAGATCACACTCACACGACCGCAGTGTGACCGCCGAGTGCAG..... 65686
710 lProGlyAspGlyLysGlyAsnLysHisLysLysHisArgLysArgArg. 726
65685CCGCACCAACGACGAGAACGACGCTGA 65658
727AsnGluGluLysGlyGluLysSerGluSerPheLeuAsnProGlu 741
65657 TCACACCGTACCGCAGACCTCGAACAACCAACGATTCGCAACCCGAC 65608
742LeuLeuGluThrSe 746
65607 CGAAACACGCTCCCGCGCTGCATTACTAGATTAACCTTCGAACACCT 65558
746 rArgLysCysArgGlySerSerGlyIleAspGluThrLysThrAspThrL 763
65557 CCGCGCGCTCTGT.....GATGAACCGCGCGCTTCGAGTCG 65523
763 euphe, ValLeuProSerArgAspAspAlaThrProValArgAspGluPr 779
65522 TGGTGATCAGCCCGCGAC.....ACGCGGTTCGCGTCCCG 65482
779 oMetAspAlaGluSerIleThrPheLys..... 788

```

645481 CAACGAAAGGGATCAATCCCGCGCTCCACGGTTTCCCAAGGACTT 65432
      ::::: ||||| ::
789 ..... SerValSerAspLysAspLysGlu 797
      ||:::||||| ::::: |||
65431 CCAGCAGCAACCCGCTGCTGGGATCCATCCGCAACCGCTCAGCGGGAA 65382
      ||||| ::::: |||
798 ... LysAspLysProLysValLysSerAspLysThrLysArgLys.... 811
      ||||| ::::: |||
65381 ATCCCGAAGAACCGGCATGCTGCGCATCTCGCAGGAAGCCACC 65332
      ||||| ::::: |||
811 ..... 811
65331 CTCACACACATAGAGGTCCCAAGACGATCGGATCCGATCCAGAACACC 65282
812 ..... SerAspLysSerAlaThrAlaLysLys 820
      ||:::||||| ::::: |||
65281 CCGCTGATCCAGCCAGCATGCTCGGAACTCCGCCACCGCATCCAG 65232
821 AspAsnValLeuLysProSer..... 827
      ||::: ||| ::
65231 CCCCCGGCCACCAACCGCAAGCTCTCCGGGAGAGACACCCACCOCGG 65182
828 ... Lys.GlyProGlnGluLysValAspLysAspArgLysSerPro 842
      ||:::||||| ::::: |||
65181 ATAAAGACAGCCCATCCCGCATCACAAGGATCTCTCAGAGTCACCG 65132
843 ArgSerGluProProLeuLys..... 849
      ||||| ::::: |||
65131 CCGCCACAGCGCGGTGCGGAGTCGAGCTTCACGCCCAACATCAGT 65082
850 ... LysAlaLysGlnGluAlaThrLysLysSerValLysProSer 865
      ||:::||||| ::::: |||
65081 TTGTGAGCAGAACCGAACCAAG.... CCGCGGTGTCGAGTATGTC 65038
865 erSerGlnLysAspGluLysValThrGlyThrProAlaLysAlaHis 881
      ||::: ||| ::::: |||
65037 GAACACCGCGAGTAGCGGCAACGAAAGCCAGACCCCGCAAGCCGT 64988
882 SerLysSerAlaLysAspThrArgArgGlnSer..... 892
      ||||| ::::: |||
64987 TCCGAATCTCCACACCGACAGCAACCAAAATCCCGGAAAGCA 64938
893 ..... GlnProArgThrArgSerLysArgThrValProLys. 905
      ||||| ::::: |||
64937 CGATCGCCCGACCTGACCGACCCGAAATGCCCAAGCAGCCCGCAAC 64888
906 ..... Thr 906
      ||||| ::::: |||
64887 ATGCTCGCGCACCAACCAACAGCTGGTTCTCTGATCCAGGTCCGCA 64838
907 SerSerGlnLysSerGlnProValArg..... ThrArgArgPro 919
      ||||| ::::: |||
64837 GCCTCGCAGCAGCAACCAACGAGCGTTGACGTCCCGCGACTCGTGTACC 64788
919 oArgSerLeuArgLysLysLeu..TyrLeuIleAlaArgGluLysAsnGlu 935
      ||||| ::::: |||
64787 CGGCTCGCGCAGCGGCGCTGGACCAAGTCCGCCAACAAGGAGGCAC 64738
936 ArgGluLysArgLys..... LysSerValAspLysAspPheGlu 949
      ||::: ||| ::::: |||
64737 ACCGCGCTCGCGCATCAGCGCGAGGGGTTTCAGGTCCAGCGTGGACG 64688
949 rSerSerMetLysLysSerLysValGluGlyThrGluIleValLysPro 966
      ||::: ||| ::::: |||
64687 GAACCAAGACGATTCGCCCGGAACCGCAACCGCATCTCGAAGACGCCG 64638
966 erProLysArgLysMetGluGlyAspValGluLysLeuGluArgThrPro 982
      ||||| ::::: |||
64637 AACCGCTCGCGCGGACATCGCGGATGCT..... 64606
983 GlnLysAspLysIleAlaSerSerThrThrProAlaLysLysIleLysLe
      ||| ::::: |||

```

```

64605 ..... GCGGCATTCGGGCTACGGCCCGCGCATCGAGC 64571
999 uAsnArgGluThrGlyLysLysIleGlyAsnAlaGluAsnAlaSerThr 1016
      ||::: ||| ::::: |||
64570 CTCGGGTCAATCCGCTGGGCTGCTCCCAAGCCCAAGCCAGCAGCACC 64521
1016 hrLysGluProSerGluLysLeuGluSerThrSerSerLysIleLysGln 1032
      ||||| ::::: |||
64520 GCAAGCAGCCC..... CGCCGTCCCGGCTGCTGAGCCAGGCCCTTCAG 64477
1033 GlnLysValLysGlyLysAla..... LysArgLysValAla 1044
      ||::: ||| ::::: |||
64476 GAACGGCTTCGCGCGCGAGTTCCTGACCCGAGCCCGCGCACGC 64427
1044 agLysSerGluGlySer..SerSerThrLeuValAspTyrThrSerThr 1060
      ||||| ::::: |||
64426 CCGAAGACGAGAGAGACGACGACGAGACGAGATCCAGATCAGAGTCC 64377
1060 rSerThrGlyLysSerProValArgLysSerGluGluLysThrAspThr 1077
      ||::: ||| ::::: |||
64376 AACTCTGACAGGTGCCAGCAGCGGCGCTTCGCGCGCAACCTTTTCG 64327
1077 yAsArgThrValIleLysThrMetGluGluTyrAsnAsnAspAsnThrAla 1093
      ||||| ::::: |||
64326 CAGGTGCTGTGGTGC..... AGCAGCTCGAGCAGC 64296
1094 ProAlaGluAspValIleIleMetIleGlnValProGlnSerLysThrAs 1110
      ||||| ::::: |||
64295 CCGTCGCGCAAGTCCCGCGCGCATGCACCACTCCGCGAGCGGAACTC 64246
1110 pLysAspAspPheGluSerGluGluGluAsp..... ValLysThr 1124
      ||||| ::::: |||
64245 TCCCGAGATCTGCTGACGACGCGGCTCAACTGGGCCGATCCCGCAT 64196
1124 hrGlnProIleGlnSerValGlyLysPro..... SerSerIleIle 1137
      ||||| ::::: |||
64195 CGCAGCTCGGCACCTGACCGCAGCTCCAGGCCACCACTGATCCACC 64146
1138 LysAsnValThr..... ThrLysProSerAlaThrAla 1148
      ||||| ::::: |||
64145 AGCTGCGTGACCCCGGGGCGCTCCACCGCGCTCCCGCCGACGACGAG 64096
1148 aLysTyrThrGluLysGluSerGluGln..... 1157
      ||||| ::::: |||
64095 CGCGCGACACCGCGCTCCCGCACCAATGCCGGGCAACCAACCGGCCCA 64046
1158 ..... ProGluLysLeuGlnLysLeuProLysGluAlaSerHis 1170
      ||||| ::::: |||
64045 GCAAGCCCGTACCGCGGAATCAACACGTTCCA..... 64011
1171 GluLeuMetGlnHisGluLeuArgSerSerLysGlySerAlaSerSerG 1187
      ||||| ::::: |||
64010 ..... TCCGACCGGCGACTCGGATCCGCT 63985
1187 uLysGlyArgAlaLysAspArgGlnHisSerGlySerGlyLysAspAsn 1204
      ||||| ::::: |||
63984 GTCTCGCGCGCAACACATCGCGCAGGAGGAGACC... CGGACTTCAC 63938
1204 roAspLysArgLysSerGlyAlaGlnProAspLys..... 1215
      ||||| ::::: |||
63937 CGCGCGCAACGCGCTGCGGTTCTCCGGAACCAACGCGCGCGCGCAGC 63888
1216 ..... GluSerThrValAspArgLeuSerGluGlnGlnHisPheLys 1229
      ||||| ::::: |||
63887 GCACGCCATATTCGGCGGTTCTCCACATCGCAACCAACGAAATCCGTC 63838
1229 sThrLeuSerGlnSerSerLysGlu..ThrArgThrSerGluLysHisGln 1245
      ||||| ::::: |||
63837 CGGTTCTCCGATCGCGCCGACGACCAACAAACCCAGAGCAGACATGCA 63788
1246 SerValArgGlySerSerAsnLysAspPheThr..... 1256
      ||||| ::::: |||
63787 CAGGTGCGAAACACCTCGCGGAGCTGTGACGCGCGACACACGATC 63738

```

```

1257 .....ProGlyArgAspLysLysValAspTyrAspSerArgAspTyrS 1271
      |||||::: :::::
63737. ACAACACACGAGGAGATCCGGGAATCGCGCCGCCCAACCAATCCTG 63688
      |||||::: :::::
1271 erSerSer.....LysArgArg 1276
      |||||::: :::::
63687 GATCAGGTCAAGCATGCGTTGTGTGCGGCGACGTGCTCGGAACACGCT 63638
      |||||::: :::::
1277 AspGluArgGlyGluLeuAlaArgArgLysAspSerProArg..... 1291
      ::: |||||::: :::::
63637 CCAAGTCTCTGGCGAGCAGCAGCGCATG.....GCACCACGACGAGCTCC 63594
      |||||::: :::::
1292 .....GlyLysGluSerLeuSerGly..... 1298
      |||||::: :::::
63593 GGGGGCGGTCAGCCCAACCGGTCAGTCCGGGTGAGATTCCGTGCCAAG 63544
      |||||::: :::::
1299 .....GlnLysSerLysLeuArgGluGluArgAspLeuProLys 1311
      ::: |||||::: :::::
63543 TCGCGAGCCCGAGGCTCGGGGAGTGTCTATGTCGATCCTCCAGTCCAGG 63494
      |||||::: :::::
1312 LysGlyAlaGluSerLysLys.....SerAsnSerSerProArg 1325
      ::: |||||::: :::::
63493 AGACCGGCTTGGCAGCCCGCGGCGCAGATTCCTTCAGTGCAGCCGGAAC 63444
      |||||::: :::::
1325 GAspLysLysProHisAspHisLysAlaProTyrGluThrLysArgProC 1342
      ::: |||||::: :::::
63443 AGGAGCTGCTGTCATCGCAC...GGGCGCTCCCAACTGCTGTCCGA 63397
      |||||::: :::::
1342 ysgGluGluThrLysProValAspLys.....AsnSer 1352
      ::: |||||::: :::::
63396 AATGTGGCGCGCACCGCATGATGATGAGACAGACGCGCTTGCCAGTCG 63347
      |||||::: :::::
1353 GlyLys GlnArgGluLysHisAlaAlaGluAlaArgAsnGlyLysGluS 1369
      ::: |||||::: :::::
63346 TGTGCGTCTAACAACCGACACCGCATTCGGCGCGCGGGAACAGAGTGC 63297
      |||||::: :::::
1369 erSerGlyAlaAsnGlyHisValTyrLeuThrArgGlnThrLeuProTyr 1385
      ::: |||||::: :::::
63296 ACCGAAGGAAGATGCGCGCTGCTTGTGAGGACACACCTTCGCACGC 63247
      |||||::: :::::
1386 ArgArg 1387
      |||||::: :::::
63246 GAACGG 63241
      |||||::: :::::
seq_name: /cogn2_6/prodata/1/lna/6B_COMB.seq:US-09-428-517-1
seq_documentation_block:
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428.517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1

```

```

alignment_scores:
  Quality: 357.50      Length: 1622
  Ratio: 0.520        Gaps: 77
  Percent Similarity: 42.355      Percent Identity: 19.975
alignment_block:
US-09-811-045a-1 x US-09-428-517-1/rev ..
Align seq 1/1 to reverse of: US-09-428-517-1 from: 1 to: 50937
96 HistHrcysProThrcysHisGlnAsnAsp..ValSerProAspAlaLeu 111
   |||||::: |||||:::
32025 CACAACCGGCCCAACATCTCAGGGCTGAGCTCAGCAGCAGCAGCAGC 31976
   |||||::: |||||:::
112 IleAlaAsnLysPheLeuArgGlnAlaValAsnAspPheLysAsn.... 126
   ::: |||||::: |||||:::
31975 GTGAGAGAGCGGCCATTGACCGCAGCCAGCAGCAGCAGCAGCAGCTC 31926
   ::: |||||::: |||||:::
127 .....GlnThrGlyThrLysArgLeuArgLysGlnLeuP 139
   |||||::: |||||:::
31925 CTCACGATGTGTGGGACACGCGCGCGGAGGCTGACGAGACATGC 31876
   |||||::: |||||:::
139 roProPheLeuPheLeuValProProProArgProLeuSerGlnArgAsn 155
   |||||::: |||||:::
31875 CGCCC.....CCACACGATCCGCGGATCCCGCGCTC 31841
   |||||::: |||||:::
156 LeuGlnProArgSerArgSerProIleLeuArgGlnGlnAspProValVa 172
   ::: |||||::: |||||:::
31840 CGCAACACACACACCGCGCGCATGCGCGCAGCTCAACCCCAACGCAC 31791
   |||||::: |||||:::
172 LpheArgTyrThrValSerProThrcysSerAspThrLysThrAlaLys 189
   ::: |||||::: |||||:::
31790 ACAAGCGGACGATCTCACCCTCGGAATGCCCAACACGACAGGCT 31741
   |||||::: |||||:::
189 er..... 189
   |||||::: |||||:::
31740 CCACACCGTAATACCGCGCGATCCGACGATGACACCATCAGCCGCAC 31691
   |||||::: |||||:::
190 .....CysSerAspSer.....GlyThrLeuSerArgLeuP 200
   |||||::: |||||:::
31690 AACGCGGCTGACCATCAACCCGCCCAAGAACAGTGGCTCCCTGCC 31641
   |||||::: |||||:::
200 Ala.....ProSerLieserLeuThrSera 210
   |||||::: |||||:::
31640 CTGCACACCTCGACACGACGACCAACCGCTCAGCGGATCGACCGCGAG 31591
   |||||::: |||||:::
31590 CGCACTCGGACACCGCTCCGCGAACCAGGACGATCCAGCAACCCA 31541
   |||||::: |||||:::
227 AlaProValProAspIleThrAlaThrValSerLieserValHisSerG1 243
   |||||::: |||||:::
31540 GCCGCCATCC.....ACCCACTGCGGA 31518
   |||||::: |||||:::
243 LysSerAspGlyProPheArgAspSerAspAsnLysLeuLeuProAlaA 260
   |||||::: |||||:::
31517 ACCCTGACACAGA.....AAGACGACACACACGAGACGACCTCAG 31477
   |||||::: |||||:::
260 IalaLeuThrSerGlnHisSerLysGlyAlaSer.SerIleAlaIleTh 276
   |||||::: |||||:::
31476 CCGGAGCCACACCGCAGACCCCGGCGCACACCCAC..... 31438
   |||||::: |||||:::
276 rAlaLeuMetGluGluLysGlyValProGlyThrSerProThrAsnSer1 293
   |||||::: |||||:::
31437 .....CCGACGCGAAGCGGCGCAACGACGACGCGCATGAGTCCA. 31399
   |||||::: |||||:::
293 LpheValGlyLysLeuLeuHisGlyGlnLeuIleProThrThrGly 309
   |||||::: |||||:::
31399 ..... 31399
   |||||::: |||||:::
310 ProValArgIleAsnAlaAlaArgProGlyGlyArgPro..GlyTyr 325
   |||||::: |||||:::

```

```

31398 ...GTTCACTGCCGATCGCAACAGCCCGGTGCTCGAAGACCGCACTCTGTC 31352
325 pgiuHisSerAsnLysLeuGlyTyrLeuValSerProProGlnGlnIleA 342
31351 G.....CCACCAATG..... 31342
342 rgarGlyGluArgSerCysTyrArgSerIleAsnArgGlyArgHisHis 338
31341 .....ACCAGCCCATCATGACCGCACTTACAC 31315
359 SerGluArgSerGlnArgThrGlnSerProSerLeuPro.....Al 372
31314 CCGACACGTCGCGCAG.....CCAAAGCTGCGCGCTTGACAGC 31277
372 atPrProCysPheValProValProProLeuTyrProPro.... 387
31276 AAGCGCCCTGCAT...CCGCGCGGACACCAACCGGACCGACAGATC 31230
388 .....ProHisThrLeuProLeuPro.....ProGlyVal 398
31229 ACCGGGACGACACAGAGTTCGCGGCTCGGCTCGGTTCCGAGGCG 31180
399 ProProGln..... 402
31179 CTTCTTCCAGATCAGATGCGCGTCTGTCGACTGACACCGAAGCGGAG 31130
403 .....PheSerProGlnPhePro..... 408
31129 ATGCGTGGCCGAGAGAGAGCGCTGCGCTCCGACGCGGCTTTC 31080
409 ..SerSerGlnProPro..... 413
31079 AGTAGCACTCCACCGCAGCCGCGGACGATCGCTCCGACGCGCTTTC 31030
414 .....ThrAlaGlyTyrSerValProProGly.. 423
31029 CATCCACATGCAAGTCTTGGGACACCAACCGCCGACGCGCAGAC 30980
424 .....PheProAlaIleAsnIle 432
30979 ATCTTCATCACACGACGACACGCGCGCTGCTATGCCGATGTT 30930
432 eThrThrLysPheSerProGlyValProThrLys..... 444
30929 CGAAGTTTACCG...ACCCAGACCAACGCGGACGCGCAGCGCCCT 30883
445 SerAsnThrMetProThrThrGlnAlaProLeuLeuSerArgGluIuph 461
30882 GCCCATACGTCGCGCAACACGCGCTGAGCCT..... 30853
461 eTyrArgGlnGlnAsnAspLysGlyArgLysPheProTyrSerG 478
30852 .....CGATCGATCACCCCAACCGCGCTCCCTAC...CATCGC 30816
478 LysSerTyrSerArgSerSerTyrThrAspSerSerGlnGlyLeuAla 494
30815 CTCACCAATCCACATCAGACGAGGACGAGCCGCTCAGCCAAACGCG 30766
495 GlnHisIleHisAlaLeuThrLeuSerProSerAlaIleHisThrLeuAs 511
30765 CAGGAGTCAACCGCTGCTGCGGAGGACCATTCGCGCGCTCAACCCATTC 30716
511 pLeuLeuHis...AspHisProHisProProGluGluAlaGluAlaArgS 527
30715 GAGCGACCATCTGATGACCGCATCCCGGACCA..... 30679
527 eValMetIleValHisMetProAspLeuMetAspIleAlaHisAla.Ar 543
30678 .....CCGCGCAACACCGCATGACCCCAACCGCGCGC 30649
543 gSerArgSerProProTyrArgArgTyr..... 552
30648 CATCCGCAACCGCTCCACCAACACACACCCACCCCTCGGACGACCA 30599

```

```

553 .....ArgSerArgSerArgSerProProGluPheArgGlyGln 565
30598 AAMCATCCGCGGAGCGGGAACGCTTGACACCGCCATCCACGACAA 30549
566 SerProThrLysArgAsnValProArgGluGluGluGluGlu..... 580
30548 ACCCG.....CTGCCGACTGAACGTCGTCACGTCACCGCGC 30511
581 .....TyrPheAsnArgTyrArgGluValProPro 591
30510 ACGACATCACCGCTCACCCACCCACCAACCGCAGGACACTACCCCGC 30461
591 rOTyrAspIleLysAlaTyrTyrGlyArgSerValAspPheArgAspPro 607
30460 CG.....CAACCC 30453
608 PheGluLysGluArgTyrArgGluTyrPGLuArgLysTyrArgGluTyrP 624
30452 CTGCACCGCCGATGCAACGCGCAACAGACGACGACGACGCGGTATCCA 30403
624 r..... 624
30402 CCGTCAACCGGAGGCCCTCAAGACCGAGCTTAAGCAACAGCCCGGAA 30353
625 .....GluLysTyrTyrLysGlyTyr... 631
30352 GCGGCCCTGATGTGTGTCGCGATGCGAGCTGCGCTTCCAGGCTTCGAC 30303
632 ..AlaValGlyAlaGlnProArgProSerAlaAsnArgLysPhe 647
30302 CAGCGGTAGGGCTGCGGCAATACCGGTCTCCACGCGCCATGTACA 30253
647 ePrProGluArgLeuLeuProLeuAsnIleArgAsnSerProPheThrArg 663
30252 CGCGCGGTGCGTGCCTTCAGAGCGGCGCGCTTGATACCGGCGCTCG 30203
664 GlyArgArgGluAspTyrAlaIleGlnSerHisArgAsnArgAsnLe 680
30202 AATGCTCTCCAGAGACGTC.....TCCAGACGACACCGCGCTCG 30165
680 uGly.....GlyAsnTyrProGluLysLeuSer 690
30164 CGGCTCATGCCAACGCTCACCGCGCGGATGCCGAAAGCGGCGCT 30115
690 hArgAspSerHisAsnAlaLysAsnProLysSerLysGluLysGlu 706
30114 CG.....AACTCCCGGAGTCACAGAAACCGCCCTCACGC 30077
707 SerGluAsnValProGly...AspGlyLysGlyAsnLysHisLysLysHis 722
30076 ACGTACAGCGTACCGCGCGCTCCGATCGGATCG.....TA 30039
722 sArgLysArgArgAsnGluGluLysGlyGluGluSerGluSerPheLeuA 739
30038 CAGGCGCCCAAGTCCACACCGCGGCTCTCGGTAGACCCGCAATGCC 29989
739 snProGluLeuLeuGluThrSerArgLysCysArgLysSerSerGlyIle 755
29988 CACGCGCGCTGTCAGCAGGTCCACAGAACTGTGCGCGGTGAGCGTG 29939
756 AspGluThrLysThrAspThrLeuPheValLeuProSerArgAspAspAl 772
29938 GCGTCCGCGCGCGCGGAGACCGGACGCTCATTCACAGTGGCGATCGG 29889
772 atPrProValArgAspGluProMetAspAlaGluSerIleThrPheLys 789
29888 CTCTCGGCGCGCTTCTCGGTCTCGGAGTCAATCGTCACTCTCTGA 29839
789 eValSerAspLysAspLysArgLysAspLysProLysValLysSer 805
29838 GTTCGTCGTGGCGCGCACAGATTCGCGCACTTCTCGGCTCACGCC 29789

```



```

28052 CTTGCCATCTGCAGACACCGCCGCTGCGCAGAGACGCGCGAG 28003
1285 ArgLysAspSerProAlaGlyLysGluSer..... 1295
28002 CGTCGACGAGTAGTCAACCGAGGCGATTCAGACAGCATGATCCGCGC 27953
1296 .....LeuSerGlyGlnLysSerLysLeuArgGlu.....G 1306
27952 CCACACATGCTCCGCGGAGACCGCTGCTGCACTCGTCTGTCGACGA 27903
1306 LuArgAspLeuProLysLysGlyAlaGluSerLysSerAsnSerSer 1322
27902 GCGGATGTGCTCTCGTCGAGGCGCTGCGCGGAGACGATCCACTTGC 27853
1323 Pro.....ProArgAs 1326
27852 CCTGTCTCGGGTGCAGACACCTCGGCATCCAGTCCGCTGCCAATGTC 27803
1326 PLYSLYSProHisAspHisLys.....AlaProTyrGluThrL 1339
27802 ACGGCGGCCATGCGCGACACCGCCGCGAGGAGTGCACACGACGACTC 27753
1339 ysArg..... 1340
27752 GCCGCGCGCGAGGCTGCGCAGTCTGTCAGGCGGTAGTACGCGGTACGA 27703
1341 ..ProCysGluGluThrLysProValAspLysAsnSerLysGluAr 1356
27702 ACGCCACCGCGACCGACGCGCGCT.....CGGCGCAG 27671
1356 gGluLysHisAlaAlaGluAlaArgAsnGlyLysGluSerSerGlyAla 1373
27670 GACCAAGCCAGCGGCGATCGTACGA.....GCATCCGGTGGTGGT 27630
1373 sn 1373
27629 CA 27628
seq_name: /cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-320-559-1
seq_documentation_block:
; Sequence 1, Application US/08320559
; Patent No. 5633135
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaan, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
; TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
; TITLE OF INVENTION: All-1 Region
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135rls
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/320,559
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/062,443
; FILING DATE: 14 MAY 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/971,094
; FILING DATE: 30-OCT-92

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,830
; FILING DATE: 27-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/805,093
; FILING DATE: 11-DEC-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TUD-0855
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1425
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ANTI-SENSE: NO
; US-08-320-559-1

```

```

alignment_scores:
Quality: 355.00 Length: 1489
Ratio: 0.590 Gaps: 64
Percent Similarity: 40.430 Percent Identity: 19.946

```

```

alignment_block:
US-09-811-045A-1 x US-08-320-559-1 ..

```

```

Align seg 1/1 to: US-08-320-559-1 from: 1 to: 14255

```

```

90 LeuLeuGluSerAspLysHisThrCysProThrCysHisGlnAsnAspVa 106
892 ATTATTCCTCTTCAAAAAGACA..... 915
106 LserProAspAlaLeuLeuAlaAsnLysPheLeuArgGlnAlaValAsn 123
916 .....GATGCAACCATTTGCTTAACAACTTACAGAGGCA..... 951
123 snPheLysAsnGluThrGlyTyrThrLysArgLeuArgLysGlnLeuPro 139
952 .....AAAAAGGGGCTCAAAAAGAAATTGAAGAAAGACGCT 990
140 ProPhe.....LeuPh 143
991 CAGCTGACGAGGAAGAAAGTGAAGACACAGTCAAAAATATTCGACAGTT 1040
143 eLeuValProProAlaProAlaProLeuSerGlnArgAsnLeuGlnProArgS 160
1041 CATCATCCCTGTTGTCAGTCTATCTCTCGCGGATCATTAAGACCCCTC 1090
160 eArgSerProLLeuArgGlnGlnAspProVal..... 172
1091 GCGCGTTTATAGAGATGAGGATATATGACCCCTCCAAATTAATGGCCGA 1140
173 .....PheArgTyrThrValSerProThrCysSerAs 183
1141 TTAGAGTCTACACCGAATAGATGATTCAGT...GCCCGTCCTGTGATC 1187
183 pThrLysThrAlaGlySerCysSerAsp.....SerLysT 195
1188 TTCTGAAAAATCAAGTGCACGCTCTCGACCTCTCTCAATATGCTTTCAG 1237
195 hLeuSerArgLeuProAlaProSerLLeuSerSerLeuThrSerAsnGln 211
1238 ACTCCTCTCGATCTAGTAGCCCGGCTTGTATCTTACCTCAGACACTCTCAG 1287
212 SerSer.....LeuAlaProProValSerGlyAsnProSe 223
1288 GCTTCTGAGGAGATTCAGTACTTCTCTGAGAGGCGGAGCATACCCCTGA 1337

```

```

223 rSerAlaProAlaProValProAspIleThrAlaThrValSerIleSerV 240
    :      ||| |||: |||
1338 AGTTCACTCT...CCACTGCC...ATTCCAGATCCC 1369
240 aHisSerGluLysSerAspGlyProPheArgAspSerAspAsnLysLeu 256
    :      :      :      :      :
1370 CAGAAATGAGATTAATGATAGGAAGACAGAAAGTATTCA..... 1410
257 LeuProAlaIleAlaLeuThrSerGluHisSerLysGlyAlaSerSerI 273
1411 .....GTGTGGAGACGAAGATTGGATCTAAGC.. 1440
273 eaIleIleThrAlaLeuMetGluLysGlyValProGlyThrSerPro 290
1440 ..... 1440
290 rPasnSerIlePheValGlyGlnSerLeuHisGlyGlnLeuIlePro 306
1440 ..... 1440
307 ThrThrGlyProValArgIleAsnAlaAlaArgProGlyGlyArgP 323
1440 ..... 1440
323 oGlyTrpGluHisSerAsnLysLeuGlyTyrLeuValSerProGlnG 340
    :      :      :      :      :
1441 .....ACGAAATAATTATCAACTCTACAAAGTCCCCCAG. 1476
340 InIleArgArgGlyGluArgSerGlyTyrArgSerIleAsnArgGlyArg 356
1476 ..... 1476
357 HisHisSerGluArgSerGluArgThrGlnSerProSerLeuProAla 373
1477 .....CAGAGACCTCTCGTCT..... 1494
373 rProCysPheValProValProProProLeuTyrProProProP 390
1495 .....CCACTCCACTCTGCTGACTCCACCGCA. 1524
390 IsthIleuProLeuProGly..... 397
1525 .....CCACTCGACCGACTCCACTATCTGACCCACACACTTGG 1566
398 ..ValProPro.....ProGlnPheSerProGlnPheProSerSerG 411
    :      :      :      :      :
1567 CTATGCTTCCACAACAATCCCTTAGCATCACATTTTGGCTTCCAC 1616
411 nProProThrAlaGlyTyrSer.....ValProProProGlyPhe. 424
    :      :      :      :      :
1617 TGCTTCCTATGCAAGGAGCGAAATCTATTTGGGAGACCGGACATTT 1666
425 .....ProProAlaProAlaAsnIleSerThr 433
1667 GGTGCACTTCTTTAAAGCATCTGAGTCAAGCCCAACTACTTTCTTCA 1716
434 AlaCysPheSer...ProGlyVal.....ProThrAlaHisSerAsn 447
    :      :      :      :      :
1717 GCAAGATGCAAGAGAGTCTTATGCAACAACAATATTGATAATTT 1766
447 rMetProThrThrGlnAlaProLeuLeuSerArgGlyGluPheTyrArg 464
    :      :      :      :      :
1767 CCGACCCCTCCACTAACTCC..... 1788
464 LuginAsnAspLysGlyArgGlnSerLysPheProTyrSerLysSer 480
    :      :      :      :      :
1789 .....GAGGACGTTGGCTTGATCTGTTTGCATCTGGTACCGCT 1833
481 TyrSerArgSerSerTyrThrAspSerSerGlnGlyLeuAlaGlnHis 497
    :      :      :      :      :
1834 GCTTACGCCGATTTTTCGSCACATCTTGACAACAAGTTGATAT 1883
497 eHisAlaLeuThrIleuSerProSerAlaAlaHisThrIleuAspLeu 514

```

```

1884 GCAC.....AAAAGAGCCCT.....CTTCTGA 1906
    :      :      :      :      :
514 IsAspHisProHisProGluGluAlaGluAlaArgSerAlaMetIle 530
    :      :      :      :      :
1907 GAGCTCCAAATTTTACTCCAAAGTAGAGCTCATCTGAAATATTGAGTCT 1956
531 ValHisMetProAspLeuMetAspIleAlaHisAlaArgSerArgSerP 547
    :      :      :      :      :
1957 GTAACCTTGCTAGTAAATCGAACTTCTGTGAACACATCTTCTCAGAGT 2006
547 oProTyrArgArgTyrArgSerArgSerArgSerProGluPheArg 564
2007 ATCCAAATGAAAAAGAAAAAGAAAGTTTATGCTTATTCGATCGATC 2056
564 LysIleSerProThrLysArgAsnValProArgGluLysGlyArgGlu 580
    :      :      :      :      :
2057 CAAGATCTCTCT...CACTCCATGAGACAAAGATGGAAGCTT 2100
581 TyrPheAsnArgTyrArgGluValProProProTyrArgIleLysAla 597
    :      :      :      :      :
2101 AGTAGTCTGAGCTTCACCTTCACCCCGCTCTGTCTCTCTC 2150
597 rTyrGlyArgSerValAsp.....P 604
2151 GTTAAAGATTTCTGTAGTCTCTTCCACTAGTCCCTTAACCCACTT 2200
604 hArgAspProPheGluLysGluArgTyrArgGluArgLysTyr 620
2201 TTACTTTCTCTCT..... 2214
621 ArgGluTrpTyrGluLysTyrTyrLysGlyTyrAlaValGlyAlaGln 637
2215 .....CAFTCCGACTCACTGTGGGAATCTGCAGAAAAATCA 2255
637 oArgPro.....SerAlaAsnArgGluAspPhe.....S 647
2256 GAGACCAAGAGACAGACTACTGCTCGGCGAGACCAATTTCATCAGATA 2305
647 eProGluArgLeuLeuProLeuAsnIleArgAsnSerProPheThrArg 663
2306 GTCTACTCTCTCTCTCCCTGGTTTACCCAGGCTCTCAGACTGAAAGA 2355
664 GlyArgArgGluAspTyrAlaAlaGlyGlnSerHisArgAsnArgAsn 680
2356 GGGAGAAATTAAGACAAAGCC..... 2376
680 uGlyLysnTyrProGluLysLeuSer...ThrArgAspSerHisAsn 696
2377 .....CCGAGGAGCTGTCCAAAGATCGAGATGCTGACAAGA 2413
696 LaLysAspAsnProLysSerLysGlyLysGluSerGluAsnValProGly 712
2414 GCGTGCAGACAGACAAAGAGTAGAGAGACGCGGAGAGA..... 2454
713 AspGlyLysGlyAsnLysHisLysLysHisArgLysAspArgAsnGlu 729
2455 ..GAAAGAGAAATAGCGGAGTCAAGAGAAAGAAAGAAAAAGG 2501
729 uLysGlyLysGluLysSerGluSerPheLeuAsnPro..... 740
2502 ATCAGAAATTCAGACTACTTCTGCTTGTATCTCTGGTAGGAGTTTCCA 2551
741 .....GluLeuLeuGluThrSerArgLysCysArgGly 751
2552 AAGAGAGGTTGTTGTAAGATGTTGCCACTTCACTTCTGCCAAAAAA 2601
752 SerSerGlyIleAspGluThrLysThrAspThrIleuPheValLeuPro 768
2602 GCACACGAGCGGAGAACAGCTTCA.....TC 2627
768 rArgAspAspAlaThrProValArgAspGluProMetAspAlaGluSer 785
    :      :      :      :      :

```

```

2628 ACATGATTCTGGGACT.....GATATTACTTCTG 2656
785 LethrheLysSerValSerAspLysArgGluLysAspLysPro 801
2657 TGACTCTT..... 2664
802 LysValLysSerAspLysThrLysArgLysSerAspLysSerAlaThrAl 818
2665 .....GGGATCAACAGC 2678
818 aLysLysAspAsnValLeu..LysProSerLysGluProGluLysVal 834
2679 TGTCAAAACCAAAATACTTATAAGAAAGAGAGAGAAATCTGAAAAA 2728
834 aLAspLysAspArgGluLysSerProArgSerGluProLeuLysLys 850
2729 CCAACTTGAC.....CTCGGCCA...ACTGCCCATCCCTGGAGAG 2769
851 AlaLysGlu.....GluAlaThrLysIleAspSerValLysProse 864
2770 GAGAAAACCTTGCTTCCTTCCACTCTTCACTAGCTGTAAACATTG 2819
864 rSerSerSer.....GlnL 869
2820 CACTCTCCTCATAGGCTCATGTTGGTCAGCAGCAACACTTCCAAATGA 2869
869 yAspGluLysValThrGlyThrProArgLysAlaHisSerLysSerAla 885
2870 CTGACACAGAGGTGGCAGCCTCTAAAAAAGCCAAAGCTCAGCTCTGC 2919
886 LysAspThrArgArgGlnSerGlnProArgThrArgArgSerLysArgTh 902
2920 AAGATTGAGAGAGTAAAGACTTTAAACAAACGACACAG..... 2958
902 rValProLysThrSerSerGlnLysSerGlnProValArgThr.....A 917
2959 ....CCAAAGCACAGGGTCACAGAAGTGACTCATCAGACCTCTGTGC 3004
917 rGArgProArgSerLeuArgLysIleAsnTyrlLeuIleAlaArgLys 933
3005 GAGGACCCCG.....ATTAAACAT...GTCTCAGAAAGACGA 3039
934 AsnGluArgGluLysArgLysSerValAspLysAspHeGluSer 950
3040 GCTGTGGCTTGCCGCGAAACGAGCTGTTCCTGAT..... 3078
950 rSerMetLysIleSerLysValGluGlyThrGluIleValLysProSer 967
3078 ..... 3078
967 rLysArgLysMetGluGlyAspValGluLysLeuGluArgThrPro... 982
3079 .....GACATGCCACCTGAGTGCCTTACCATGG 3108
983 ...GluLysAspLysIleAlaSerSerThrThrProAlaLysLysIle 998
3109 GAAGAAGAGAGAAAGATTGTCTTC..... 3135
998 sLeuAsnArgGluThrGlyLysIleGlyAsnAlaGluAsnAlaSer 1015
3136 .ATGGGGAATGATGACAGATCATCAATGCTGCTCGAAGATGCTGAC 3184
1015 hrThrLysGluProSerGluLysLeuGluSer...ThrSerSerLysIle 1030
3185 CTCTGTCTCCACCCATCAACCAATTAAACCTGTCACTAGAAACAGCA 3234
1031 LysGluGluLys.....ValLysGlyAlaLysArgLysValAlaGlu 1045
3235 CCCGAGAACCTCCAGTAAGAAAGACGTCGATCGAGCGGTGGGCA 3284
1045 ySerGluGly.....SerSerSerThrLeuValAspTyr 1057
3285 GTGTCCCGGCTGCCAGGTGCTGAGGACTGTGTGTGTACTAATTCGT 3334

```

```

1057 hrSerThrSerSerThrGlyLysSerProValArgLysSerGluGluLys 1073
3335 TAGATAACCCAAAGTTGTGGTGCATATTAAGAAAGAGCTGTCAG 3384
1074 ThrAspThrLysArgThrValIleLysThrMetGluGluTyrAsnAsn 1090
3385 .....ATGAGAAAATGTCAGAAATCT 3404
1090 rAsnThrAlaProAlaGluAspValIleIleMetIleGlnValProGln 1107
3405 ACAATGATGCTCTCCAAACCTACCTGAGAGCAAGCTAAAGCTGTGA 3454
1107 eLysThrAspLysAspAspPheGluSerGluGluGluLysValLysThr 1123
3455 AAAAGAAAGAGAAAGTCTAAGACCAAGTAAAGAAAGAC..... 3495
1124 ThrGlnProIleGlnSerValGlyLysProSerSerIleIleLysAsn 1140
3496 .....ACCAAGAGACGAGTGTGTGAAGAACGT 3524
1140 LhrThrLysProSerAlaThrAlaLysTyrThrGluLysGluSerGlu 1157
3525 GGTG.....GACTTACGTACGA 3541
1157 LhrProGluLysLeuGlnLysLeuProLysGluAlaSerHisGluLeu 1173
3542 AACCT..... 3546
1174 GlnHisGluLeuArgSerSerLysGlySerAlaSerSerGluLysGly 1190
3546 ..... 3546
1190 gAlaLysAspArgGluHisSerGlySerGluLysAspAsnProAspLys 1207
3547 .....ACCCATCAGCAAGAGAGATCTGCCCCAA 3577
1207 rGlySerGlyAlaGlnProAspLysGluSerThrValAspArgLeuSer 1223
3578 AGAAAGCAGTACTGAGCCTCCACAGAAAGCCGTCGAGGAAAGAGT 3627
1224 GluGlnGlyHisPheLysThrLeuSerGlnSerSerLysGluThrArg 1240
3628 GAGAGAGGATGTCTCGGCCCTGGGCTGATTCCAACAGGCCACAC 3677
3678 TCCAGCTTCCAGAGTCAAGCAAGCAGCTCTCCAGCCAGCACTGCTCA 3727
1256 .....ThrProGlyArgAspLysLysValAspTyrAsp 1266
3728 TCCCGCTCAGCCACCTACTACAGAGCCGCCAAGAAA..... 3765
1267 SerArgAspTyrSerSerSerLysArgArgAspGluArgGlyGluLeu 1283
3766 .....CAAGTTC 3773
1283 aArgArgLysAspSerProProArgGlyLysGluSerLeuSerGlyGln 1300
3774 CAAAACACACTCTAGTGCAGCCCAAGAAAGACAGCT..... 3810
1300 ySerLysLeuArgGluGluArgAspLeuProLysLysGlyAlaGluSer 1316
3811 .....CCACCACCAAAATGAGTCCAGCAGCAG 3837
1317 LysLysSerAsnSerSerProProArgAspLysLysProHisAspHis 1333
3838 AGCAAAACAGAAAAAGTGGCTCCCGCCCAAGTATCCCTTAACAAAA 3887
1333 sAlaProTyrGluThrLysArgProCysGluGluThrLysProValAsp 1350
3888 A...CCAAAGAAAGAGGAAAGCA.....CTCCGCTCAATA 3922

```

```
1350 YS..... 1350
1351 .. 1351
1352 ACCAGAGAGATGACGACACTTGAACATCCTGACGACTCTCCATGAC 1352
1353 AenSerglyLysGluArg..... 1356
1354 .. 1354
1355 AATAGTCTTAACGCAAAATATCCAGAGATGAGATCCACAGATGAGACT 1355
1356 .. 1356
1357 .. 1357
1358 .. 1358
1359 .. 1359
1360 .. 1360
1361 .. 1361
1362 .. 1362
1363 .. 1363
1364 .. 1364
1365 .. 1365
1366 .. 1366
1367 .. 1367
1368 .. 1368
1369 .. 1369
1370 .. 1370
1371 .. 1371
1372 .. 1372
1373 .. 1373
1374 .. 1374
1375 .. 1375
1376 .. 1376
1377 .. 1377
1378 .. 1378
1379 .. 1379
1380 .. 1380
1381 .. 1381
1382 .. 1382
1383 .. 1383
1384 .. 1384
1385 .. 1385
1386 .. 1386
1387 .. 1387
1388 .. 1388
1389 .. 1389
1390 .. 1390
1391 .. 1391
1392 .. 1392
1393 .. 1393
1394 .. 1394
1395 .. 1395
1396 .. 1396
1397 .. 1397
1398 .. 1398
1399 .. 1399
1400 .. 1400
1401 .. 1401
1402 .. 1402
1403 .. 1403
1404 .. 1404
1405 .. 1405
1406 .. 1406
1407 .. 1407
1408 .. 1408
1409 .. 1409
1410 .. 1410
1411 .. 1411
1412 .. 1412
1413 .. 1413
1414 .. 1414
1415 .. 1415
1416 .. 1416
1417 .. 1417
1418 .. 1418
1419 .. 1419
1420 .. 1420
1421 .. 1421
1422 .. 1422
1423 .. 1423
1424 .. 1424
1425 .. 1425
1426 .. 1426
1427 .. 1427
1428 .. 1428
1429 .. 1429
1430 .. 1430
1431 .. 1431
1432 .. 1432
1433 .. 1433
1434 .. 1434
1435 .. 1435
1436 .. 1436
1437 .. 1437
1438 .. 1438
1439 .. 1439
1440 .. 1440
1441 .. 1441
1442 .. 1442
1443 .. 1443
1444 .. 1444
1445 .. 1445
1446 .. 1446
1447 .. 1447
1448 .. 1448
1449 .. 1449
1450 .. 1450
1451 .. 1451
1452 .. 1452
1453 .. 1453
1454 .. 1454
1455 .. 1455
1456 .. 1456
1457 .. 1457
1458 .. 1458
1459 .. 1459
1460 .. 1460
1461 .. 1461
1462 .. 1462
1463 .. 1463
1464 .. 1464
1465 .. 1465
1466 .. 1466
1467 .. 1467
1468 .. 1468
1469 .. 1469
1470 .. 1470
1471 .. 1471
1472 .. 1472
1473 .. 1473
1474 .. 1474
1475 .. 1475
1476 .. 1476
1477 .. 1477
1478 .. 1478
1479 .. 1479
1480 .. 1480
1481 .. 1481
1482 .. 1482
1483 .. 1483
1484 .. 1484
1485 .. 1485
1486 .. 1486
1487 .. 1487
1488 .. 1488
1489 .. 1489
1490 .. 1490
1491 .. 1491
1492 .. 1492
1493 .. 1493
1494 .. 1494
1495 .. 1495
1496 .. 1496
1497 .. 1497
1498 .. 1498
1499 .. 1499
1500 .. 1500
```

seq_name: /cgr2_6/ptodata/1/ina/5A_COMB.seq:US-08-327-392-1

seq_documentation_block:

Sequence 1, Application US/08327392

Patent No. 5633136

GENERAL INFORMATION:

APPLICANT: Croce, Carlo

APPLICANT: Canaan, Eli

TITLE OF INVENTION: ALL-1 Polynucleotides and Monoclonal

TITLE OF INVENTION: Antibodies for Leukemia Detection and

TITLE OF INVENTION: Treatment

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633136ris

STREET: One Liberty Place - 46th Floor

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE

COMPUTER: IBM

OPERATING SYSTEM: PC-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/327,392

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/971,094

FILING DATE: 30-OCT-92

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/888,830

FILING DATE: 27-MAY-92

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/805,093

FILING DATE: 11-DEC-91

ATTORNEY/AGENT INFORMATION:

NAME: Deluca, Mark

REGISTRATION NUMBER: 33,229

REFERENCE/DOCKET NUMBER: TJU-1331

TELEPHONE: (215) 568-3100

TELEFAX: (215) 568-3439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 14255
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ANTI-SENSE: No
US-08-327-392-1

alignment_scores:
Quality: 355.00 Length: 1489
Ratio: 0.590 Gaps: 64
Percent Similarity: 40.430 Percent Identity: 19.946

alignment_block:
US-09-811-045A-1 x US-08-327-392-1 ..

Align seg 1/1 to: US-08-327-392-1 from: 1 to: 14255

```
90 LeuLeuGluSerAspLysHisThrCysProThrCysHisGlnAsnAspVal 106
107 .. 107
108 .. 108
109 .. 109
110 .. 110
111 .. 111
112 .. 112
113 .. 113
114 .. 114
115 .. 115
116 .. 116
117 .. 117
118 .. 118
119 .. 119
120 .. 120
121 .. 121
122 .. 122
123 .. 123
124 .. 124
125 .. 125
126 .. 126
127 .. 127
128 .. 128
129 .. 129
130 .. 130
131 .. 131
132 .. 132
133 .. 133
134 .. 134
135 .. 135
136 .. 136
137 .. 137
138 .. 138
139 .. 139
140 .. 140
141 .. 141
142 .. 142
143 .. 143
144 .. 144
145 .. 145
146 .. 146
147 .. 147
148 .. 148
149 .. 149
150 .. 150
151 .. 151
152 .. 152
153 .. 153
154 .. 154
155 .. 155
156 .. 156
157 .. 157
158 .. 158
159 .. 159
160 .. 160
161 .. 161
162 .. 162
163 .. 163
164 .. 164
165 .. 165
166 .. 166
167 .. 167
168 .. 168
169 .. 169
170 .. 170
171 .. 171
172 .. 172
173 .. 173
174 .. 174
175 .. 175
176 .. 176
177 .. 177
178 .. 178
179 .. 179
180 .. 180
181 .. 181
182 .. 182
183 .. 183
184 .. 184
185 .. 185
186 .. 186
187 .. 187
188 .. 188
189 .. 189
190 .. 190
191 .. 191
192 .. 192
193 .. 193
194 .. 194
195 .. 195
196 .. 196
197 .. 197
198 .. 198
199 .. 199
200 .. 200
201 .. 201
202 .. 202
203 .. 203
204 .. 204
205 .. 205
206 .. 206
207 .. 207
208 .. 208
209 .. 209
210 .. 210
211 .. 211
212 .. 212
213 .. 213
214 .. 214
215 .. 215
216 .. 216
217 .. 217
218 .. 218
219 .. 219
220 .. 220
221 .. 221
222 .. 222
223 .. 223
224 .. 224
225 .. 225
226 .. 226
227 .. 227
228 .. 228
229 .. 229
230 .. 230
231 .. 231
232 .. 232
233 .. 233
234 .. 234
235 .. 235
236 .. 236
237 .. 237
238 .. 238
239 .. 239
240 .. 240
241 .. 241
242 .. 242
243 .. 243
244 .. 244
245 .. 245
246 .. 246
247 .. 247
248 .. 248
249 .. 249
250 .. 250
251 .. 251
252 .. 252
253 .. 253
254 .. 254
255 .. 255
256 .. 256
257 .. 257
258 .. 258
259 .. 259
260 .. 260
261 .. 261
262 .. 262
263 .. 263
264 .. 264
265 .. 265
266 .. 266
267 .. 267
268 .. 268
269 .. 269
270 .. 270
271 .. 271
272 .. 272
273 .. 273
274 .. 274
275 .. 275
276 .. 276
277 .. 277
278 .. 278
279 .. 279
280 .. 280
281 .. 281
282 .. 282
283 .. 283
284 .. 284
285 .. 285
286 .. 286
287 .. 287
288 .. 288
289 .. 289
290 .. 290
291 .. 291
292 .. 292
293 .. 293
294 .. 294
295 .. 295
296 .. 296
297 .. 297
298 .. 298
299 .. 299
300 .. 300
```

```

1440 ..... 1440
290 rpaSerSerIlePheValGlyInSerLeuHisGlyInIlePro 306
1440 ..... 1440
307 ThrThrGlyProValArgIleAsnAlaIaArgProGlyGlyArgPr 323
1440 ..... 1440
323 oGlyTrpGlnHisSerAsnLysLeuGlyTrpLeuValSerProGlnG 340
1441 .....ACGAAAAATTATCACTGTACAAAGTCCCCCAG. 1476
340 InIleArgGlyGluArgSerCysTrpArgSerIleAsnArgGlyArg 356
1476 ..... 1476
357 HisHisSerGluArgSerGlnArgThrGlnSerProSerLeuProAlaTh 373
1477 .....CAGAGACCTCCTGCTCT..... 1494
373 rProCysPheValProValProProProLeuTrpProProProH 390
1495 .....CCACCTCCACCTGTGTGATCCACGCCCA. 1524
390 IsthrLeuProLeuProProGly..... 397
1525 .....CCAGTCGACCGACCTCCAGTATCTGTGACACACACCTTGG 1566
398 ...ValProPro.....ProGlnPheSerProGlnPheProSerSerG 411
1567 CTATAGCTCCCAACAATCCCTTATAGCATCACTATTTGGCTGCTTCCAC 1616
411 nProProThAlaGlyTrpSer.....ValProProGlyPhe. 424
1617 TGCCTATGCAAGGAAAGCAAAATCTATTGCGAAGCCGACATTTA 1666
425 .....ProProAlaProAlaAsnIleSerThr 433
1667 GGTGGACTTTTAAAGCATTTAGTGTACAGCCACAATCTTTCTCTCA 1716
434 AlacysPheSer...ProGlyVal.....ProThrAlaHisSerAsnTh 447
1717 GCAAGATATGCCAAGAAGAGCTTATTCGCAACCAATATTGTGATTTT 1766
447 rMetProThrThrGlnAlaProLeuLeuSerArgGlyGluPheTrpArg 464
1767 CCGACCCCTCCACTACTCC..... 1788
464 LuginAsnAspLysGlyArgGluSerLysPheProTrpSerGlySerSer 480
1789 .....GAGGACGTTGGCTTGCATCTGTTTTCTGTGATCTGTGACGCT 1833
481 TyrSerArgSerSerTrpThrAspSerSerGlnGlyLeuAlaGlnHisI 497
1834 GCTTCACCCGAGATTGTTTGGCCACCTCCATTCGTGAAACAAGGTTGAT 1883
497 eHisAlaLeuThrLeuSerProSerAlaAlaHisThrLeuAspLeuH 514
1884 GCAC.....AAAAGAGCCCT.....CTTCTGA 1906
514 IsAspHisProHisProGlnGluAlaGluAlaArgSerAlaMetIle 530
1907 GAGCTCCACAATTTATCTCAAGTAGGCTCACTGTAGAATATTTGAGCT 1956
531 ValHisMetProAspLeuMetAspIleAlaHisAlaArgSerArgSerP 547
1957 GTAACCTTGCTAGTATCGAAGCTTCGTGGAACATCTTCTTCAGAGT 2006
547 oProTrpArgTrpArgSerArgSerArgSerProProGluPheArgG 564
2007 ATCCAAATAGAAAAAGAAAAAGTGTATTAGTCTATTCTGATCTGAAC 2056

```

```

564 LysInSerProThrLysArgAsnValProArgGluGlyLysGluArgGlu 580
2057 CAAGATCTCCTCT.....CAGTCCATGAGACCAAGAAAGTGAAGGCTT 2100
581 TyrPheAsnArgTrpArgGluValProProProTrpAspIleLysAlaTy 597
2101 AGTAGTTCGAGCTTCACCTCTCACCCCTCCACCCCTCTCTCTCTCTC 2150
597 rTrpGlyArgSerValAsp.....P 604
2151 GTTAAGCATTTCTGTATGCTCTCTGACACAGTACCTTAACCAACTT 2200
604 heArgAspProPheGlnLysGluArgTrpArgGluTrpGluArgLysTr 620
2201 TTACTTTTCTCTCT..... 2214
621 ArgGluTrpTrpGluLysTrpTrpLysGlyTrpAlaValGlyAlaGlnP 637
2215 .....CATTCCTGACTCAGCTGTGGGGAATCTGCAGAGAAAAATCA 2255
637 oArgPro.....SerAlaAsnArgGluAspPhe.....S 647
2256 GAGACCAAGAGACAGACAGTGTGCTCCGACAGACCATTTTCATCAAGTA 2305
647 eTrpGluArgLeuLeuProLeuAsnIleArgAsnSerProPheThrArg 663
2306 GTCTCTCTCTCTCTCTCTCTGTTTACCCAGGCTCTCAGACTGAAGA 2355
664 GlYArgArgGluAspTrpAlaAlaGlyInSerHisArgAsnArgAsnLe 680
2356 GGGAGAAATTAAGACAGGCC..... 2376
680 uGlyLysAsnTrpProGluLysLeuSer...ThrArgAspSerHisAsnA 696
2377 .....CCCGAGAGCTGTCCAAAGATCGAGATGTGACACAGA 2413
696 lAlysAspAsnProLysSerLysGluLysGluSerGluAsnValProGly 712
2414 GCGTGGAAGAGACACAGATGAGAGAGAGACCGGAGAGA..... 2454
713 AspGlyLysGlyAsnLysHisLysLysHisArgLysArgArgAsnGlu 729
2455 ...GAAAGAGAGATATAGCGGAGTCAAGGAAAGAAAGAAAAAGGG 2501
729 uLysGlyGluLysSerGluSerPheLeuAsnPro..... 740
2502 ATCAGAAATTACAGAGTAGTCTGCTTGTATCCTGTGGTAGGCTTCCA 2551
741 .....GluLeuLeuGluThrSerArgLysCysArgGly 751
2552 AAGCAAGAGTGTGTTGTAAGATGTGCCACTTCATGTTCTGCCAAAAA 2601
752 SerSerGlyIleAspGluThrLysThrAspThrLeuPheValLeuPro 768
2602 GCAACAGGGGGAGAGAGTCTCA.....TC 2627
768 rArgAspAspAlaThrProValArgAspGluPrometAspAlaGluSerI 785
2628 ACATGATTCGTGGACT.....GATATTACTTCTGT 2656
785 lerThrPheLysSerValSerAspLysAspLysArgGluLysAspLysPro 801
2657 TGACTCTT..... 2664
802 LysValLysSerAspLysThrLysArgLysSerAspLysSerAlaThrAl 818
2665 .....GGGATACAAACAGC 2678
818 aLysLysAspAsnValLeu...LysProSerLysGlyProGlnGluLysV 834
2679 TGTCAAAACCAAAATACTTATTAAGAAAGGAGAGGAAATCTGTGAAAAA 2728

```



```

373 rProcySpheValProValProProProLeuTyProProProProH 390
      |||||
1495 .....CCACCTCCACCTCTCTGACCTCCACCGCCA. 1524
390 lsrHleuProLeuProProGly..... 397
      |||||
1525 .....CCACTGCAGCCAGCTCCAGTATCTGCACACACACTTGG 1566
398 ...ValProPro.....ProGlnPheSerProGlnPheProSerSergl 411
      :|||
1567 CTATGCTCCACCAACATCCCTTAGATACACCATTTTTCCTGCTCCAC 1616
411 nProProThrAlaGlyTySer.....ValProProProGlyPhe. 424
      |||
1617 TGCCTCATGACAGGAGGAAATATATTTCGAGAACCGCATTTA 1666
425 .....ProProAlaProAlaAsnIleSerThr 433
      |||
1667 GGTGACTCTTTTAAAGCATTTCTAGTCAAGCCACAAATACTTTTCTCA 1716
434 AlaCysPheSer...ProGlyVal.....ProThrAlaHisSerAsnTh 447
      |||
1717 GCAAAGATATGCCAAGAGGCTTATTTCGCAAAACCAATATTGTAATTT 1766
447 rMeProThrThrGlnAlaProLeuLeuSerArgGluIuPheTyArg 464
      |||
1767 CCGACCCCTCCACCTACTCC..... 1788
464 lueIAsnAspPlyGlyArgGluSerLysPheProTySerglySerSerg 480
      |||
1789 ....GAGGAGGTTGGCTTGATCTGTTTTCGATCTGGTACCGCT 1833
481 TySerArgSerSerTyThrAspSerSergGlnGlyLeuAlaGlnHisIL 497
      |||
1834 GCTTACGCCGATTTGTTTCGCCACTCATTTGGAACAAGTTGATAT 1883
497 eHisAlaLeuThrLeuSerProSerAlaAlaHisThrLeuAspLeuH 514
      |||
1884 GCAC.....AAAGAGCCCT.....CTCTGA 1906
514 lAspHisProHisProProGluGluAlaGluAlaArgSerAlaMetIle 530
      |||
1907 GAGCTCCAAAGATTACTCTCAAGTGAGGCTCCTAGAAATATTGAGCTCT 1956
531 ValHisMetProAspLeuMetAspIleAlaHisAlaArgSerArgSerPr 547
      |||
1957 GTTACCTTGCTAGTATATCGAATCTTGCTGGAACATCTTCTTCAGAGCT 2006
547 oProTyArgArgTyArgSerArgSerArgSerProProGluPheArg 564
      |||
2007 ATCCAATAGAAAAAGAAAGAAAGTGTAGTCTATTCGATCTGAC 2056
564 lylGlnSerProThrlLysArgAsnValProArgGluGluLysGluArgGlu 580
      |||
2057 CAAGATCTCTCTCT.....CACTCCATGAGACAAAGAGGAGGCTT 2100
581 TyrPheAsnArgTyArgGluValProProProTyArgSpleLysAlaTy 597
      |||
2101 AGTAGTTAGCTCTACCTCTACCCCGCTCTCTCTCTCTCTC 2150
597 rTyGlyArgSerValAsp.....P 604
      |||
2151 GTTAAAGCATTTCTGTAGTCTCTTGGCAGTATGCTTAAACCAACTT 2200
604 heArgAspProPheGluLysGluArgTyArgGluIuPheGluArgLysTy 620
      |||
2201 TACTTTCTCTCT..... 2214
621 ArgGluIuPheTyArgGluLysTyArgGluLysGluTyArgAlaValGluArgInPr 637
      |||
2215 .....CATTCCTGACTCACTGCTGGGGAATCTGCAGAGAAAAATCA 2255

```

```

637 oArgPro.....SerAlaAsnArgGluAspPhe.....S 647
      |||||
2256 GAGCCCAAGAGAGAGACTAGTGTCCGGCAGAGCCATTTTCAATCAGTA 2305
647 erProGluIuArgLeuProLeuAsnIleArgAsnSerProPheThrArg 663
      |||||
2306 GTCTACTCTCTCTCTCTGTTTACCAGCCAGCTCTCAGACTGAAAGA 2355
664 GlYArgArgGluAspTyArgAlaGlyGlnSerHisArgAsnArgAsnLe 680
      |||||
2356 GGGAGAAATAAAGCAAGGCC..... 2376
680 uGlyGlyAsnTyProGluLysLeuSer...ThrArgAspSerHisAsnA 696
      |||||
2377 .....CCGAGAGCTGTCCAAAGATCGAGATGCTGACAAAGA 2413
696 lAlYsAspAsnProLysSerLysGluLysGluSerGluAsnValProGly 712
      |||
2414 GCGTGGAGAAAGGACAAAGATGAGAGAGAGACCGGAGAGA..... 2454
713 AspGlyLysGlyAsnLysHisLysLysHisArgGlySerArgAsnGluI 729
      |||
2455 ...GAAGAAGAGATTAAGCGGAGCTCAAGAAAGAGAAAGAAAGAGG 2501
729 uLysGlyGluGluSerGluSerPheLeuAsnPro..... 740
      |||
2502 ATCAGAAATTCAGAGTAGTCTGCTTGTATCCGTGGTAGGTTTCCA 2551
741 .....GluLeuGluIuThrSerArgLysCysArgGly 751
2552 AAGAGAGGTTGTTGGTGAAGAGTGTCCACTTCATCTTGCACAAAATA 2601
752 SerSerglyLileAspGluIuThrLysThrAspThrLeuPheValLeuPro 768
      |||
2602 GCACAGCGCGGGAAGAGCTTCA.....TC 2627
768 rArgAspAspAlaThrProValArgAspGluPrometAspAlaGluSerI 785
      |||
2628 ACATGATTTCTGGGACT.....GATATTACTTCTG 2656
785 lenthPheLysSerValSerAspLysAspLysArgGluLysAspLysPro 801
      |||
2657 TGACTCTT..... 2664
802 lysValLysSerAspLysThrLysArgLysSerAspLysSerAlaThrAl 818
      |||
2665 .....GGGATACACAGC 2678
818 alYsLysAspAsnValLeu...LysProSerLysGlyProGlnGluLysV 834
      |||
2679 TGTCAAACCAACCAATACTTATAAGAAAGGAGAGGAAATCTGGAAAAA 2728
834 alAspGlyAspArgGluLysSerProArgSerGluProProLeuLys 850
      |||
2729 CCAACTTGGAC.....CTGGGCCA...ACTGCCCATCTCTGGGAAG 2769
851 AlaLysGlu.....GluAlaThrLysIleAspSerValLysProse 864
      |||
2770 GAGAAACCCCTCTGCTTCCACCTCTTCAATCTAGCACTGTAAACATTC 2819
864 rSerSerSer.....GlnL 869
      |||
2820 CACTTCTCCATAGGCTCCATGTTGGCTCAGAGACACAACTTCCAAATGA 2869
869 ysAspGluLysValThrGlyThrProArgLysAlaHisSerLysSerAla 885
      |||
2870 CTGACAAAGAGGTTGCGAGCTCTTAATAAAGGCCAAAGCTCAGCTGTC 2919
886 LysAspThrArgArgGlnSerGlnProArgThrArgArgSerLysArgTh 902
      |||
2920 AAGTTTGAAGAAGATAGAGTCTTAAACAAACGACAGC..... 2958
902 rValProLysThrSerSergLysSerGlnProValArgThr.....A 917

```



```

2959 .....CCCAAGCAGGCTCAGAAAGTCACTATCAGAGACCTCTGTGC 3004
917 rgargproargserleuarglysileasntyrlleuilealaarglulys 933
3005 GAGAGACCCCGG.....ATTAACAT...GTCTCAGAGAAGCA 3039
934 Asngluaraglulysarglyserservalasplysasphegluserse 950
3040 GCGTTCCTCCCTGGCCGAAAACGAGCTGTCTTCCTGAT..... 3078
950 rsermetlysileserlysvalgluglythrgluilevallysproserp 967
3078 ..... 3078
967 rplysarglysmetgluglyaspyalglulysleugluargthrpro... 982
3079 .....GACATGCCACCTGAGTGCCTTACCATCG 3108
983 ...glulysasplysileaserthrthrproalalsylsilely 998
3109 GAGAGACGAGAAAAGATTGTCTTC..... 3135
998 sleuasnaargluthrlygllyslysilegluasnaalaglualasert 1015
3136 .ATGGGAATGATGACACAGTCATCAATTGCTGCTCAGAAATGCTGAAC 3184
1015 hrthrlysgluoproserglulysleugluser...thrserlysile 1030
3185 CTCTGTCTCCACCCATCAACCAATTAACTGTCTCCTAGAACCAAGCA 3224
1031 lysglnglulys.....Vallysglyalalsarglysvalalagl 1045
3235 CCCAGGAACTCCAGTAAGAAAGACGTGATCGAGCGGTGGGCA 3284
1045 yserglugly.....SerSerSerThrleuvalaspyrtyr 1057
3285 GTGTCCCGCTGCCAGGTGCTGAGACTGTGTGTGTGTAATTAATGCT 3334
1057 hrserthrserSerThrlygllyserprovalarglyserserglulys 1073
3335 TACATAGCCCAAGTTGTGTGTGCGCAATATATAAGAACAGTGTGCAAG 3384
1074 Thrasphrlysrargthrvalilelysthrmetgluglutyriasnana 1090
3385 .....ATGAGAAAATGTCAGAATCT 3404
1090 pasnthrAlaProAlaGluaspvalileilemetilegluvalproglins 1107
3405 ACATATGATGCTTCCAAAGCCTACCTGCAAGAACAGCAAACTGTGA 3454
1107 erlystrpasplysaspasphegluserglugluaspvalylsThr 1123
3455 AAAAGAAAGGAAAAAGCTAAGCACGTGAAAAAGAAAGAC..... 3495
1124 ThrGluProIleGlnserValGlyLysProSerSerIlelelysasna 1140
3496 .....AGCAAAAGAGAGCAGTGTGTGGAAGAACGT 3524
1140 lthrthrlysproseralathralalyslyrThrGluLysGluSerGlu 1157
3525 GGTG.....GACCTGTGTGACA 3541
1157 lnrpogluLysleuGluLysleuProLysglualaserHisGluLeu 1173
3542 AACCT..... 3546
1174 GluHisGluLeuArgSerSerLysGlySerAlaSerSerGluLysGlyAr 1190
3546 ..... 3546
1190 galalysasparggluHisserGlySerGluLysAspAspProaspLysA 1207
3547 .....ACCCATCAGCAAGAGAGATCTGCCCAA 3577
1207 rglYserSerGlyAlaGlnProaspLysGluSerThrValAspArgLeuSer 1223
3578 AGAAAGACAGTATGAGACTCTCTCCACGAAAGCCCGTCGAGAAAAGAGT 3627
1224 GluGlnGlyHisPheLysThrLeuSerGlnSerSerLysGluThrArgTh 1240
3628 GAAAGAGGAATGCTCTGCGCCCTGGGCTGATCCAAACAGGCCACAC 3677
1240 rSerGluLysHisGluSerValArgGlySerSerAsnLysAspHe... 1255
3678 TCCAGTTCAGGAAGTCAAGCAAGCAGGTCTCCACAGCAGACTGTGTCA 3727
1256 .....ThrProGlyArgaspLysLysValAspTyrrasp 1266
3728 TCCCGCTCAGCCACCTACTACAGAGCCGCAAGAAAA..... 3765
1267 SerArgAspTyrrSerSerSerLysArgArgaspLuarglugluLeuAl 1283
3766 .....GAGTTC 3773
1283 aargarglyaspsrProProarglyLysGluSerLeuSerGlyGlnL 1300
3774 CAAGAACACTCTCTAGTGAGCCCAAGAAAAAGCAGCT..... 3810
1300 yserlysleuargglugluargspLeuProLysLysGlyAlaGluSer 1316
3811 .....CCACCACCAAGATCAGGTCCAGACAG 3837
1317 LysLysSerAsnSerSerProProargAspLysLysProHisPhtsly 1333
3838 AGCAAAAGAAAAAAGTGCGTCCCGCCCAAGTATCCCTGTAAAAACAAA 3887
1333 salaprotYrGluThrlysrargProCysGluGluThrLysProValaspl 1350
3888 A...CCAAAGAAAAAGAAAAACCA.....CTTCGGGTCAATA 3922
1350 ys..... 1350
3923 AGCAGGAAATGCAGGACCTTGAACATCTCAGCAGCTCTCTCCAAATGC 3972
1351 AsnserGlyLysGluArg..... 1356
3973 AATAGTCTTAAGCAAAAAATTCACGACAGATGAGATCCACAGATCAAGT 4022
1357 ...GluLysHisAlaAlaGluAlaArgAsnGlyLysGluSerSerGlyA 1372
4023 GCACTTTAAGAGGATGTGTGAAGCAGAAATGTGTGGAGATGGAGCGCT 4072
1372 laasn..... 1373
4073 TAGGAATCTTGACTTCTGTCTCTATAACACCAGCGGTGTGTCTCTC 4122
1374 .....CysHisVal..... 1376
4123 TGTCACAGTATGAGGCAATGTAAGACTTTGTGATTTGCCAAGTCTTGTGA 4172
1377 .....TyrlleuThrArgGlnThrLeuPro...TrpArgArg 1387
4173 GCCCTTCACAAAGTTTGTGTAGAGAGAACGACGCCCTGTGAGAGACC 4222
1388 SerTrpLeuLeuGly 1392
4223 AGCTGAAAAATGTGT 4237
seq_name: /cogn2_6/prodata/1/ina/6A_COMB.seq:us-08-545-860D-1
seq_documentation_block:
; Sequence 1, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
```

APPLICANT: Canaan, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
ADDRESS: No. 6040140ris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,860D
FILING DATE: 07-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE: 22-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10930
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/327,392
FILING DATE: 19-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/320,559
FILING DATE: 11-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,443
FILING DATE: 14-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,094
FILING DATE: 30-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/888,839
FILING DATE: 27-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,093
FILING DATE: 11-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Deluca Esq., Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-1262
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 14255
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ANTI-SENSE: No
US-08-545-860D-1
alignment_scores:
Quality: 355.00 Length: 1489
Ratio: 0.590 Gaps: 64
Percent Similarity: 40.430 Percent Identity: 19.946
alignment_block:
US-09-811-045A-1 x US-08-545-860D-1 ..
Align seg 1/1 to: US-08-545-860D-1 from: 1 to: 14255

```
90 LeuLeuGluSerAspIysHisThrCysProThrCysHisGlnAsnAspVal 106
      ::::: |||:::|||||
892 ATTATTCCTTCTCAAAAGACA..... 915
106 LserProAspAlaLeuIleAlaAsnLysPheLeuArgGlnAlaValAsnA 123
      ||||| |||||:::|||||
916 .....GATGCACCATTCCTTAGCAACCTTAGAGAGGCA..... 951
123 snPheLysAsnGluThrGlyTyrThrLysArgGluLeuArgLysGlnLeuPro 139
      ::: || |||:::|||||
952 .....AAAAGGGGCTCAAAAGAAATGTGAAAAGAGCAGCT 990
140 ProPhe.....LeuPh 143
      :::
991 CAGCTGACGGAGAAAGTGAAAGACACAGCTCAAAATATTCGACAGCT 1040
143 eleValProProProArgProLeuSerGlnArgAsnLeuGlnProArgS 160
      :::||||| :::|||||
1041 CATCATCCCTGTGTGAGTGCTATCTCCGCGGATCATAGACCCCTC 1090
160 eArgSerProIleLeuArgGlnAsnProValVal..... 172
      ||| ::: |||||
1091 GCGCGTTATAGAGAGATATGACCCCTCAATTAATTTGCCGA 1140
173 .....PheArgTyrThrValSerProThrCysSerAs 183
      |||::: |||:::
1141 TTAGAGCTCTACCCGAATATAGATTAGT..GCCCGCTCTGTGATC 1187
183 pThrLysThrAlaGlySerCysSerAsp.....SerLysT 195
      ::: |||::: |||:::
1188 TTCTGAAATAATCAAGTCAGCTTCTCAGACCTCCTCAATGCTTCAG 1237
195 hrlUserArgLeuProAlaProSerIleSerSerLeuThrSerAsnGln 211
      ||||| ::: |||||
1238 ACTCCTCTCGATCTAGTACCCCGAGTGTGATCTCCACAGACTCTCAG 1287
212 SerSer.....LeuAlaProProValSerGlyAsnProSe 223
      ::: ||| ||| ::: |||
1288 GCTTCTGAGGAGATTCAGTACTCTCTGAGAGGAGCGATACCCCTGA 1337
223 rSerAlaProAlaProValProAspIleThrAlaThrValSerIleSerV 240
      ||| |||::: |||
1338 AGTTCATCCT...CCACTGCC.....ATTTCACAGTCC 1369
240 alHisSerGlnLysSerAspIleProPheArgAspSerAsnLysLeu 256
      ::: |||::: |||
1370 CAGAAATGAGATGATGATAGAGAACGACAGATATCA..... 1410
257 LeuProAlaAlaAlaLeuThrSerGlnHisSerLysGlyAlaSerSerI 273
      ::: |||||::: |||
1411 .....GTGCGAGAGAGTTTGGATCTAGAACG.. 1440
273 eAlaIleThrAlaLeuMetGluGlnLysGlyValProGlyThrSerProT 290
1440 ..... 1440
290 rPasnSerIlePheValGlyGlnSerLeuLeuHisGlyLnuLeuIlePro 306
1440 ..... 1440
307 ThrThrGlyProValArgIleAsnAlaAlaArgProGlyGlyArgPr 323
1440 ..... 1440
323 oGlyTyrPglHisSerAsnLysLeuGlyTyrLeuValSerProPogIng 340
      ::: |||||::: |||
1441 .....ACGAAATAATATCAACTCTACAAAGTCCGCCCCAG. 1476
340 InlIeArgArgGlyGlnArgSerCysTyrArgSerIleAsnArgLysArg 356
1476 ..... 1476
357 HisHisSerGluArgSerGlnArgThrGlnSerProSerLeuProAlaTh 373
```

1477	: :	..CAGGAGACCTCTCTGCTT	1494
373	rProCysPheValProValProProProLeuTyrProProProProH		390	
1495CCACCTCCACCTCTGCTGACTCCACCGCCA	1524	
390	1stThrLeuProLeuProGly	397	
1525CCACTGCAGCCAGCCCTCAGATCTGTGACCAACACACTTGG	1566	
398	...ValProPro.....	ProLinPheSerProGlnPheProSerSerG1	411		
1567	CTTATGCTCCACCAACAATCCCTTTAGCATCCACATTTTGGCTGTCTCAC	1616		
411	nProProThrAlaGlyTyrSer.....	ValProProProGlyPhe	424		
1617	TGCTCTTATGCAGGAGGAGCAAAATCTATTTTGGGAGAACCGACATTTA	1666		
425ProProAlaProAlaAsnIleSerThr	433		
1667	GGTGAGACTCTTAAAGCAATCTGTAGTGCAGACCCACAAATACTTTCTCTCA	1716		
434	AlaCysPheSer...ProGlyVal.....	ProThrAlaHisSerAsnTh	447		
1717	GCAAACTATGCCAAMAGAAAGTCTGTATTGCCAAMCAATATTGTGAATAATT	1766		
447	rMetProThrThrGlnAlaProLeuLeuSerArgLysIlePheTyrArg	464			
1767	CGGAGCCCTCCACATACCTCC.....	1788		
464	IuGlnAsnAspLysGlyArgGluSerLysPheProTyrSerGlySerSer	480			
1789	...GAGACGTTGGCTTGTGCATCTGTTTTTTCGCATCTGGTACCGCT	1833			
481	TyrSerArgSerSerTyrThrAspSerSerGlnGlyLeuAlaGlnHisI1	497			
1834	GCYTTCAGCCGATGTTTTCGCACATCTGTGGAAACAAGGTGTGATAT	1883			
497	eHisAlaLeuThrLeuSerProSerAlaAlaHisThrLeuAspLeuLysH	514			
1884	GCAC.....AAAAGAGCCCT.....	1906		
514	IsaSPHisProHisProProGluGlnAlaGluAlaArgSerAlaMetIle	530			
1907	GACCTCCAGATTTACTCCAAAGTAGAGCCCACTCAGAAATATTAGCT	1956			
531	ValHisMetProAspLeuMetAspIleAlaHisAlaArgSerArgSerPr	547			
1957	GTAACTCTGCCTGATAGTACAACTTCTGTGGAAATCTTCTTCAGAGT	2008			
547	oProTyrArgArgTyrArgSerArgSerArgSerProProGluPheArg	564			
2007	ATCCAAATAGAAAAAGAAAAAGAAAGATGTTTGATCTCTATTGCATCTGAA	2056			
564	LysIleSerProThrLysArgAsnValProArgLysIleLysGluArgLys	580			
2057	CAGAGATCTCTCTT.....	CACCTCATGAGACAGAAAGTGAAGGCTT	2100		
581	TyrPheAsnArgTyrArgLysValProProProTyrAspIleLysAlaTyr	597			
2101	AGTAGTTCGTGAGCTGCACCTCTGCACCCCGCGTCTCTGTCTCTCTGC	2150			
597	rTyrGlyArgSerValAsp.....P	604		
2151	GTTAACCACTTCTGTGTAAGCTCTTGCACATGAGCTTAAACCCAACTT	2200			
604	heArgAspProPheGluLysGluArgTyrArgGluTyrProLysArgLysTyr	620			
2201	TTACTTTTCTCTT.....	2216		
621	ArgGluTyrPheGluLysTyrTyrLysGlyTyrAlaValGlyAlaGlnPr	637			

2215CATTCCTGACTAGTCTGGGGAATCTGCAGAGAAAAATCA 2235

637 oArgPro.....SerAlaAsnArgLysAspHe.....S 647

2256 GAGCCCAAGGAAGCAGACGTAAGTGTCCGGCAGACGCCATTTTCATCAAGTA 2305

647 eProGluArgLeuLeuProLeuAsnLLeuArgAsnSerProPheThrArg 663

2306 GTCTCTACTCTCTCTCTCCCTGGTTTACCCTCCAGGCTCTCAGACTGMAAGA 2355

664 GLVArgArgLysGluSeryrLalaIaLglnSerHisArgAsnArgAsnLe 680

2356 GGGAGAAATTAAGACAAGGCC..... 2376

680 uGlyGlyAsnTyProGluLysLeuSer...ThrArgAspSerHisAsnA 696

2377CCCGAGAGCTGTCCAAAGATGAGATGCTGTGCACAGA 2413

696 IalysAspAsnProLysSerLysGluLysGluSerGluAsnValProGly 712

2414 GCGTGGAGAAGCAGACAGACTAAGAGAGAGACCGGAGAGA..... 2454

713 AspGlyLysGluLysAsnLysHisLysHisArgLysArgArgAsnGlu 729

2455 ...GAAAGAGAGATTAAGCGGGGAGTCAAGAAAGAAAGAAAAAGG 2501

729 uLysGlyGluLysSerGluSerPheLeuAsnPro..... 740

2502 ATCAGAAATTCAGAGTAGTCTGTGCTTTGTATCTCTGGGTAGGGTTCCA 2551

741GluLeuGluThrSerArgLysCysArgLys 751

2552 AAGAGAAAGTGTGTGGTGAAGATGTGCCACTTCATCTTCCAAAAA 2601

752 SerSerGlyLLeuSpcLuthrLysThrAspThrLeuPheValLeuProse 768

2602 GCACACAGGGCGGAAGAAGTCTCA.....TC 2627

768 rArgAspAspAlaThrProValArgAspLuprometAspLagLuserI 785

2628 ACATGATCTGGGACT.....CATTTACTTCTG 2656

785 IeThrPheLysSerValSerAspLysAspLysArgLulysAspLysPro 801

2657 TGACTCTT..... 2664

802 LysValLysSerAspLysThrLysArgLysSerAspGlySerAlaThrAl 818

2665GGGAGATCAACAAGC 2678

818 aLysLysAspAsnValLeu...LysProSerLysGlyProGlnGluLysV 834

2679 TGTCAAAACCAAAATCTTATTAAGAAAGGAGAGAAATCTGCAAAAAA 2728

834 aLAspArgLysArgGluLysSerProArgSerGluProProLeuLysLys 850

2729 CCAACTGTGAGC...CTCGGCCA...ACGCCCCATCCTGTGAGAG 2769

851 ALaLysGlu.....GluAlaThrLysLLeuAspSerValLysProse 864

2770 GAGAAACCTCTGCTTCACACTCTTCATGTACACACTGTAAACATTC 2819

864 rSerSerSer.....GlnL 869

2820 CACTCTCTCATAGGCTCCATGTTGGCTCAGCAGACAAGCTTCCAATGA 2869

869 yAspArgLysValThrGlyThrProArgLysAlaHisLysSerLysAla 885

2870 CTGCAAGAGAGGTTGCCAGCTCTCTAAAAAGCCAAAGCTAGCTGCG 2919

886 LysAspThrArgArgLnserGlnProArgThrArgArgSerLysArgTh 902

2920 AAGTTTGAAGTAAGTAAGTCTTAAACCAACCGGACAG..... 2958

APPLICANT: Crocco, Carlo
APPLICANT: Cnaan, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities In the All-1-1
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
ADDRESS: Morris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca Esq., Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJD-1242
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 14255
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ANTI-SENSE: NO
PCT-US94-04496-1

```
alignment_scores:
  Quality: 355.00
  Ratio: 0.590
  Gaps: 64
Percent Similarity: 40.430
Percent Identity: 19.946
```

```
alignment_block:
US-09-811-045A-1 x PCT-US94-04496-1 ..
```

Align seg 1/1 to: PCT-US94-04496-1 from: 1 to: 14255

```

90 LeuLIGLuserAspLysHisIshrcYsyrProthrcYsHisLsInAsnAspVa 106
   ::::: |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
892 ATTATTCCCTCTTCAAAAAGGCA..... 915

106 lserProaspAlaLeu||leAlaAsnLysPheLeuArgInAlaIvalAsna 123
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
916 .....GATGCACCATTTGCTTCAACCACTCTTACAGAGGCA..... 951

123 snPheLysasnLunthrGlyTyrThrLysArgLeuArgLysGlnLeuPro 139
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
952 .....AAAAAGGGGCTCAAAAAGAAATTGAAAAAGAACACAGCT 990

140 ProPhe.....LeuPh 143
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
991 CAGCTGCACGGAAGAAGAGTGAAAGACACAGTCAAAAATATTGACAGATT 1040

143 eleuValProProArgProLeuSerGlnArgAsnLeuInProArgS 160
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1041 CATCATGCTGTTGTCAGAGTCATATCTCCTCGGGAATCATTTAAGACCCCTC 10900

160 eArgSerProIleLeuArgGlnAspProVal..... 172
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1091 GGCGGGTTATAGAGATGAGTGGATTTTGTGAACCCCTCCAAATTAAATGCCCGA 11400

```

[illegible]


```

3109 GAAGACGAGAAAAAGATTGCTCTCC..... 3135
998 steuasnargluThrGlyLysLysIleGlnAsnAlaGluAsnAlaSerT 1015
      :::::|||||:.....:|||||:
3136 .ATGGGAAATGATGACACAGTCAATCAATGCTGCTCAGAAAGATGTCGAC 3184
1015 hThrLysGluProSerGluLysLeuGluSer...ThrSerLysIle 1030
      |||:.....:|||||:
3185 CTCTTCTCTCCACCACATCAACCAATTAACCTGCTCACTAGAAACAGGCA 3234
1031 LysGlnGluLys.....ValLysGlyLysAlaLysArgLysValAlaG 1045
      |||||:.....:|||||:
3235 CCCACAGAACCTCCAGTAAAGAAAGACGTCAGCGGGGTGTGGGCA 3284
1045 ySerGluGly.....SerSerThrLeuValAlaSPtyrT 1057
      |||:.....:|||||:
3285 GTGTCCCGGCTGCCAGTGCCTGAGACGTGTGTGTTGACTTAATGTCT 3334
1057 hSerThrSerSerThrGlyLysSerProValArgLysSerGluLys 1073
      :::::|||||:.....:|||||:
3335 TAGATAGCCCAAGTTGGTGTGTCGCAATTTAAAGACAGTCTGCAAG 3384
1074 ThrAspThrLysArgThrValIleLysThrMetGluGluTyrAsnAs 1090
      |||:.....:|||||:
3385 .....ATGAGAAAATGTCAGATCT 3404
1090 pAsnThrAlaProAlaGluAspValIleIleMetIleGlnValProGln 1107
      :::::|||||:.....:|||||:
3405 ACATGAGATGCTTCCAAAGCTTACTCAGAGCAAGCAAGCTAAAGCTGTA 3454
1107 eLysTrpAspLysAspAspPheGluSerGluGluAspValLysThr 1123
      ::|||:.....:|||||:
3455 AAAAGAAAGAAAAGTCTAAGACCAAGTAAAGAAAGAC..... 3495
1124 ThrGlnProIleGlnSerValGlyLysProSerSerIleLysAsnVa 1140
      ::|||:.....:|||||:
3496 .....ACCAAGAGACAGTGTGTGAAGACGT 3524
1140 LThrThrLysProSerAlaThrAlaLysTyrThrGluLysGluSerGlu 1157
      |||:.....:|||||:
3525 GGTG.....GACTTATGTCAGA 3541
1157 LProGluLysLeuGlnLysLeuProLysGluAlaSerHisGluLeuMet 1173
      ::|||:
3542 AACCT..... 3546
1174 GlnHisGluLeuArgSerSerLysGlySerAlaSerSerGluLysGlyAr 1190
3546 ..... 3546
1190 gAlaLysAspArgGluHisSerGlySerGluLysAspAsnProAspLysA 1207
3547 .....ACCCATCAGCAAGAGAGATCCGCCCAA 3577
1207 rGlySerGlyAlaGlnProAspLysGluSerThrValAspArgLeuSer 1223
      ::|||:.....:|||||:
3578 AGAAAGACAGTACTGAGCTCTCCACGAAAGCCCTCGAGGAAAGACGT 3627
1224 GluGlnGlyHisPheLysThrLeuSerGlnSerSerLysGluThrArgTh 1240
      |||:.....:|||||:
3628 GAAGAAGGAGATGTCTCGGCCCTCGATCCAAACAGGCCACAC 3677
1240 rSerGluLysHisGluSerValArgGlySerSerAsnLysAspPhe... 1255
      |||:.....:|||||:
3678 TCCAGCTTCAGAGAGTCAAGCAAGCTCTCCAGCCGACACTGTGTCA 3727
1256 .....ThrProGlyArgAspLysValAspTyrAsp 1266
      |||:
3728 TCCCGCTCAGCCACTACTACAGACCCCAAGAAAA..... 3765
1267 SerArgAspTyrSerSerSerLysArgArgAspGluArgLysGluLeuAl 1283
      |||:
3766 .....GAAGTTCC 3773

```

```

1283 aArgArgLysAspSerProProArgGlyLysGluSerLeuSerGlyGlnL 1300
      :::::|||||:.....:|||||:
3774 CAAMACCACTCTCTAGTGAGCCCAAGAAAAAGACGCT..... 3810
1300 ySerLysLeuArgGluGluArgAspLeuProLysLysGlyAlaGluSer 1316
      |||:.....:|||||:
3811 .....CCACACCAAGATCAGTCCAGACGAC 3837
1317 LysLysSerAsnSerSerProProArgAspLysLysProHisAspHis 1333
      ::|||:.....:|||||:
3838 AGCAAAACAGAAAAAGTGGCTCCCGCCCAAGTATCCCTTAACAAAAA 3887
1333 sAlaProTyrGluThrLysArgProCysGluGluGluThrLysProValAspL 1350
      |||:.....:|||||:
3888 A...CCAAAAAGAAAAAACCA.....CTCCGGTCAATA 3922
1350 ys..... 1350
3923 AGCAGAGAAATGACGACCTTTGAACATCTCTCAGCACTCTCTCAATGGC 3972
1351 AsnSerGlyLysGluArg..... 1356
3973 AATAGTTCTTACGCAAAAAATTCCAGCAGATGAGTCCACAGATCAGAGT 4022
1357 ....GluLysHisAlaAlaGluAlaArgAsnGlyLysGluSerSerGlyA 1372
4023 GCACTTTAAGAGGATGTGTAACAGAAAAATGTGTGGAGATGGAGGCT 4072
1372 laSn..... 1373
4073 TAGGAATCTTGACTTCTGTTCTTCTAATAACCCAGGAGGTTGCTTTCTC 4122
1374 .....CysHisVal..... 1376
4123 TGTGCCAGTAGTGGCATGTAGAGTTGTGTATTTGCCAAGTCTGTGTGA 4172
1377 .....TyrLeuThrArgGlnThrLeuPro.TrrPArgArg 1387
4173 GCCCTTCCACAAAGTTTGTTTAAGAGGAGAAAGAGAGCCCTCTGGAGAGCC 4222
1388 SerTrpLeuLeuGly 1392
      |||||:
4223 AGCTGGAATTTGCT 4237
seq_name: /cgn2.6/ptodata/1/ina/6B.COMB.seq:US-08-061-376-4
seq_documentation_block:
; Sequence 4, Application US/08061376
; Patent No. 6175000
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; APPLICANT: Dyabali, Malek
; APPLICANT: Selleri, Lucia
; APPLICANT: Parry, Pauline
; TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11023
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/061,376
; FILING DATE: 13-MAY-1993

```

```

: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Reiter, Stephen E.
: REGISTRATION NUMBER: 31,192
: REFERENCE/DOCKET NUMBER: P41 9387
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619)546-4737
: TELEFAX: (619)546-9392
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 11907 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
: US-08-061-376-4

alignment_scores:
  Quality: 345.50      Length: 1568
  Ratio: 0.537         Gaps: 72
  Percent Similarity: 41.008   Percent Identity: 20.026

alignment_block:
  us-09-811-045a-1 x US-08-061-376-4  ..

Align seg 1/1  to: US-08-061-376-4  from: 1  to: 11907

18  GLYLSYRAlaIleProThrIleAspAlaGlu.AlaTyraIaIleGlyL 34
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
876  GGAAGGAAGGGGTCAAAATTGTAGACGGAGAGGAAGGCCCTCCATCAACA 925
   34  ySLySGlULySProPhe..LeuPro.....GluGluP 45
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
926  GAAAGATTAAGACCCCTTCGCTCCATATTCTGAACTGGAAAGC 975
   45  rSeSerSeSerGluGluAspProIleProAlaGluLeuLeuCyS 61
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
976  CCCAAGAAAGTCGGAAAGACAAGAACACCTCA..... 1013
   62  LeuIleCySLySAspIleMethrAspAlaValIleProCyCySGL 78
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1014  ..CTTACAAAAGAA.....GATAGACAGTTGTC..... 1040
   78  yAsnSeSerCySAspGluCySLeaGThrThrLeuLeuGluSerAspL 95
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1041  .AGACAAAGCCCTCGAAGGATTAAAGCCAGTTAGATTCTCTTCA 1089
   95  yShStHcySProThrcyShStGlnAsnAspValSerProAspAlaLeu 111
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1090  AAAGGACA.....GATCCAAAC 1106
   112  IleAlaAsnLySPhelLeuArgIleAlaValAsnAsnPhelYsAsnGluTh 128
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1107  ATTGCTAAGCAACTCTTACAGAGGCA.....AAAA 1138
   128  rGlyTyThr..LysArgLeuArgLySGLnLeuProPhe..... 141
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1139  AGGGGCTCAAAAGAAATTTGAAAAAGAACGCTCAGCTCAGCGAAGA 1188
   142  .....LeuPheLeuValProProPr 148
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1189  AAGGTGAAGACACAGGTCAAAAATATTCGACAGATCATCATGCGTGTGT 1238
   148  cArgProLeuSerGlnArgAsnLeuGlnProArgSerArgSerProIleL 165
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1239  CAGTCTATCTCTCGCGATCATTAAGACCCCTCGCGGTTTATTAAGAG 1288
   165  euArgGlnGlnAspProValVal..... 172
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1289  ATGAGATTATAGACCTCCAAATTAAATTTGCCGATTAGACTCAACCG 1338
   173  ..PheArgTyThrThValSerProThrcySerAspThrLysThAlaGlu 188
```

```

1339  AATAGTATTCAGT...GCCCGTCTGTGGATCTTGAAATAATCAAG 1385
   |||:|||||  |||:|||||  |||:|||||  |||:|||||
168  ySerCySserAsp.....SerGlyThrLeuSerArgLeuP 200
   |||:|||||  |||:|||||  |||:|||||  |||:|||||
1386  TGCACCTTCTGCACCTCTCAAAATGCTTCACAGACTCTTCATCTA 1435
   200  rAlaProSerIleSerSerLeuThrSerAsnGlnSer..... 213
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1436  GTAGCCCAAGTGTATATACCTCCACAGACTCTCAGGCTTCAGAGATT 1485
   214  .....LeuAlaProProValSerGlyAsnProSerSerAlaProAlaPr 228
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1486  CAGTACTTCTCGAGACGCGAGCATACCCCTGAAGTTCACTT...CC 1532
   228  oValProAspIleThrAlaThValSerIleSerValHisSerGlyYSS 245
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1533  ACTGCC.....ATTCCAGTCCCAAGAAATGAGAGTA 1567
   245  eAspSGlyProPheArgAspSerAspAsnLySLeuLeuProAlaAla 261
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1568  ATGATAGGAGAAAGCAGAGGTATTCA..... 1593
   262  LeuThrSerGlnHisSerLySGlyAlaSerSerIleAlaIleThAlaLe 278
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1594  ..GTTCGAGAGAGAAATTGTGATCTAGAACG..... 1623
   278  uMetGluGlySLySGlyValProGlyThrSerProThrAsnSerIlePheV 295
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1623  ..... 1623
   295  aLgLyGlnSerLeuLeuHisGlyGlnLeuIleProThrGlyProVal 311
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1623  ..... 1623
   312  ArgIleAsnAlaAlaArgProGlyGlyGlyArgProGlyTyThrGlnHisSe 328
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1624  .....AC 1625
   328  rAsnLySLeuGlyTyLeuValSerProProGlnGlnIleArgArgLyS 345
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1626  GAAATAATTATCACTTACAAAGTGCCCCCAGAG..... 1662
   345  LuArgSerCySLyGArgSerIleAsnArgLyArgHisHisSerGluArg 361
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1662  ..... 1662
   362  SerGlnArgThrGlnSerProSerLeuProAlaThrProCySPheValPr 378
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1663  ..CAGACCTCTCTGCTCA..... 1680
   378  oValProProProProLeuTyTyProProPro..... 389
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1681  ...CTCCACCTCTGCTGACTCCACGCCCACTGACAGCCAGCTCCA 1726
   390  .....HisThrLeuPro.....LeuProProGlyValProPro 400
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1727  GTATCTTGACACACA...CTTGGCTTATGCGCCCAACATC..... 1767
   401  ProGlnPheSerProGlnPheProSerSerGlnProProThrAlaGlyTy 417
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1768  CCTTAGCATCACCTTTTGGCTGCTTCCACTGCTCTTATGCAAGGAA 1817
   417  rSer.....ValProProGlyPhe..... 424
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1818  GCGAAATCTATTTTGGAGAAAGCAGACATTTAGTGAGACTTTTAAAGC 1867
   425  ....ProProAlaProAlaAsnIleSerThrAlaCySPheSer...Pro 438
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1868  ATTCTAGTCAAGAGCCACAATATCTTCCACAGCAAGATATGCAAGAA 1917
   439  GlyVal.....ProThrAlaHisSerAsnThrMetProThrGlnAla 453
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
```



```

1918 GGTCTATTGCAAAACAATATTGATTTCCAGCCCTCCACTAAC 1967
453 aProLeuLeuSerArgGluGluPheTyrArgGluGlnAsnAspLysGlyA 470
1968 TCCC.....GAGGACGTTGGCT 1984
470 rGluSerLysPheProTyrSerGlySerSerTyrSerArgSerSerTyr 486
1985 TTGCATCTGGTTTTCGATCTGTGACCGCTGCTTACGCCCATGTTT 2034
487 ThrAspSerSerGlnGlyLeuAlaGlnHisIleHisAlaLeuThrLeuSe 503
2035 TCGCCACTTCATCTGGAACAAGTTTGATATGCAC.....AAAGAG 2078
503 rProSerAlaAlaHisThrLeuAspLeuLeuHisAspHisProHisPro 520
2079 CCTT.....CTTCTGAGAGCTCCACAATTTATCTC 2107
520 rGluGluAlaGluAlaArgSerAlaMetIleValHisMetProAspLeu 536
2108 CAAGTCAGGCTCACTCTAGATATTGAGTCTGTACCTTGCTTACTAAT 2157
537 MetAspIleAlaHisAlaArgSerArgSerProProTyrArgArgTyrAr 553
2158 CGAACTTCTGTGGAACATCTTCTCAGAGATGCCAATAGAAAAGAA 2207
553 gSerArgSerArgSerProProGluPheArgGlyGlnSerProThrLysA 570
2208 AAGAAAAGTTTATGCTATTCGATCTGACCAACCAAGATCTCTTCT... 2253
570 rGlnAsnValProArgGluGluGlyLysGluArgGluTyrPheAsnArgTyrArg 586
2254 ..CACCTCCATGAGACACAGACAGAGGCTTAGTGTCTGAGCTCTCA 2301
587 GluValProProProTyrAspIleLysAlaTyrTyrGlyArgSerValAs 603
2302 CCTCTACCCGCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2351
603 P.....PheArgAspProPheGluL 610
2352 TCCTCTTGCCACTAGTGCCTTAAACCACTTTTCTCTCTCT... 2397
610 yGluLysArgTyrArgGluTyrGluArgLysTyrArgGluTyrPryArgLys 626
2398 .....CATTCCTCG 2406
627 TyrTyrLysGlyTyrAlaValGlyAlaGlnProArgPro..... 639
2407 ACTCACTCTGGGGAATCTGCAGAGAAAATTCAGACCAAGAGAGAGAC 2456
640 .SerAlaAsnArgGluAspPhe.....SerProGluArgLeuLeuP 653
2457 TAGTGTCTCCGCGAGACCATTTTCATCAAGTAGTACTCTCTCTCC 2506
653 rLeuLeuAsnIleArgAsnSerProPheThrArgLysArgArgLysAspTyr 669
2507 CTGTGTTTACCCAGGCTCTCAGACTGAAAGAGGAGAAATTAAGACAA 2556
670 AlaAlaGlyIleSerHisArgAsnArgAsnLeuGlyLysAsnTyrProG 686
2557 GCC.....CCCGA 2564
686 uLysLeuSer...ThrArgAspSerHisAsnAlaLysAspAsnProLys 702
2565 GGAGCTGTCCAAAGATCGAGATGCTGACAAAGACGTGAGAGAGAGACA 2614
702 rLysGluLysGluSerGluAsnValProGlyAspGlyLysGlyAsnLys 718
2615 GTAGAGAGAGACCGGAGAGAG.....GAAAGGAGAAATTAAG 2652
719 HisLysLysHisArgLysArgArgAsnGluGluLysGlyGluGluSerG 735
2653 CGGAGTCAAGAAAGAAAGAAAGGAGATCAAGAAATTCAGAGTAG 2702

```

```

735 uSerPheLeuAsnPro.....G 741
2703 TTCTGCTTTGTATCTGTGGTAGGCTTCCAAAGAGGTTGTGGTG 2752
741 IuLeuLeuGluThrSerArgLysCysArgGlySerSerGlyIleAspGlu 757
2753 AAGATTTGCACTTCTCATCTGTCCCAAAAAGCAACAGCGCGAAGAA 2802
758 ThrLysThrAspThrLeuPheValLeuProSerArgAspAspAlaThrP 774
2803 TCTTCA.....TCACATGATCTGCGGACT... 2826
774 oValArgAspGluProMetAspAlaGluSerIleThrPheLysSerVal 791
2827 .....GATATTACTTCTGTGACTCTT..... 2847
791 eAspLysAspLysArgGluLysAspLysProLysValLysSerAspLys 807
2847 ..... 2847
808 ThrLysArgLysSerAspGlySerAlaThrAlaLysLysAspAsnVal 824
2848 .....GGGATACACACAGCTGTCAAAAACCAAAATACT 2879
824 u...LysProSerLysGlyProGlnGluLysValAspGlyAspArgGlu 840
2880 TATTAAGAAAGGAGAGAAATCTGAAAAAACCACTTGAC.....C 2923
840 ySerProArgSerGluProProLeuLysLysAlaLysGlu..... 853
2924 TCGGCCCA...ACTGCCCATCCCTGAGAGAGAGAAAACCTTGCTT 2970
854 GluAlaThrLysIleAspSerValLysProSerSerSer..... 867
2971 TCCACCTCTCATCTGACACTGTTAAACAATTCACACTCTCCATAGGCTC 3020
868 .....GlnLysAspGluLysValThrG 875
3021 CATGTTGGCTCAGGACAGACAGCTTCCAAATGACTGACACAGAGGTTGCCA 3070
875 LysThrProArgLysAlaHisSerLysSerAlaLysAspThrArgArgGln 891
3071 GCCTCTTAAAAAAGGCCAAAGCTCAGCTCTGCAAGATGTGAAGAGTAAG 3120
892 SerGlnProArgThrArgArgSerLysArgThrValProLysThrSerSe 908
3121 AGCTTAAACAAACGACAGC.....CCCAAGCACAGGG 3155
908 rGlnLysSerGlnProValArgThr....ArgArgProArgSerLeuA 923
3156 TCAAGAAAGAGACTCATCAGAGACCTGTGCGAGAGACCCCGG..... 3198
923 rGlyLysIleAsnTyrLeuIleAlaArgGluLysAsnGluArgGluLysArg 939
3199 .....ATTAAACAT...GTCTGCAGAAAGACACTGTGCTTGCCCGCA 3240
940 LysLysSerValAspLysAspPheGluSerSerSerMetLysIleSerLys 956
3241 AAACGAGCTGTGTTCCCTGAT..... 3261
956 sValGluGlyThrGluIleValLysProSerProLysArgLysMetGluG 973
3261 ..... 3261
973 LysAspValGluLysLeuGluArgThrPro.....GluLysAspLysIle 987
3262 ..GACATGCCCACTGAGTGCCTTACCATGGAAGACAGAGAAAGAT 3309
988 AlaSerSerThrThrProAlaLysLysIleLysLeuAsnArgGluThrG 1004
3310 TTGCTCTCC.....ATGGCAATGATGACAA 3335

```

```

1004 yLysIleGlyAsnAlaGluAsnAlaSerThrThrLysGluProSerG 1021
      ::::::::::::::::::::|::|
3336 GTCAATTCCTGCTGCTGAGATGATCTGATCTGCTCCACCATCA 3385
      ::::::::::::::::::::|::|
1021 LuLysLeuGluSer...ThrSerLysIleLysGlnLys... 1034
      ::::::::::::::::::::|::|
3366 AACCAATTAAACCTGCTCACTAGAACAGCCACGACGACCTCCAGTA 3435
      ::::::::::::::::::::|::|
1035 ValLysGlyLysAlaLysArgLysValAlaGlySerGluGly... 1048
      ::::::::::::::::::::|::|
3436 AAGAAAGACGTCGATCGAGCGGCTGTGGCAGTGTCCCGCTCCAGGT 3485
      ::::::::::::::::::::|::|
1049 .....SerSerSerThrLeuValAspTyrThrSerThrSerSerThrG 1063
      ::::::::::::::::::::|::|
3486 GCCTGAGCTGTGCTGTTTACTAATTCCTTATGATTAAGCCCAAGTTTG 3535
      ::::::::::::::::::::|::|
1063 LyGlySerProValArgLysSerGluGluLysThrAspThrLysArgThr 1079
      ::::::::::::::::::::|::|
3536 GTGGTCGCAATTAAGAAAGAGCTGCTGCAG... 3567
      ::::::::::::::::::::|::|
1080 ValIleLysThrMetGluGluTyrAsnAsnAspAsnThrAlaProAlaG 1096
      ::::::::::::::::::::|::|
3568 .....ATGAGAAATGTCAGAAATCAATGATGATGCTTCCAA 3605
      ::::::::::::::::::::|::|
1096 uAspValIleIleMetIleGlnValProGlnSerLysTrpAspLysAspA 1113
      ::::::::::::::::::::|::|
3606 AGCCTACCTGCAGAAAGCAAGCTAAGCTGTGAAAAAGAAAGAAAAAGT 3655
      ::::::::::::::::::::|::|
1113 sPheGluSerGluGluGluAspValLysThrThrGlnProIleGlnSer 1129
      ::::::::::::::::::::|::|
3656 CTAAAGACGATGAAAAGAAAGAC... 3678
      ::::::::::::::::::::|::|
1130 ValGlyLysProSerSerIleIleLysAsnValThrThrLysProSerAl 1146
      ::::::::::::::::::::|::|
3679 ...ACCAAGAGAGCAGGTGTGTGAAGAACGTGTG... 3711
      ::::::::::::::::::::|::|
1146 aThrAlaLysTyrThrGluLysGluSerGluGlnProGluLysLeuGlnL 1163
      ::::::::::::::::::::|::|
3712 .....GACTCTAGTCAGAACCT... 3729
      ::::::::::::::::::::|::|
1163 yLeuProLysGluAlaSerHisGluLeuMetGlnHisGluLeuArgSer 1179
      ::::::::::::::::::::|::|
3729 ..... 3729
      ::::::::::::::::::::|::|
1180 SerLysGlySerAlaSerSerGluLysGlyArgAlaLysAspArgGluH 1196
      ::::::::::::::::::::|::|
3729 ..... 3729
      ::::::::::::::::::::|::|
1196 sSerGlySerGluLysAspAsnProAspLysArgLysSerGlyAlaGlnP 1213
      ::::::::::::::::::::|::|
3730 .ACCCATTCAGCAAGAGAGATCTGCCCAAAAGAAAGCACTAGTGAGC 3778
      ::::::::::::::::::::|::|
1213 roAspLysGluSerThrValAspArgLeuSerGluGlnGlyHisPheLys 1229
      ::::::::::::::::::::|::|
3779 CTCCTCCACGAAAGCCGTCGAGGAAAGAGTGAAGAGGAAATGTCG 3828
      ::::::::::::::::::::|::|
1230 ThrLeuSerGlnSerSerLysGluThrArgThrSerGluLysHisGluSe 1246
      ::::::::::::::::::::|::|
3829 GCCCCTGGGCTGAAATCCAAACAGCCACACTCCAGCTTCAGAGATGC 3878
      ::::::::::::::::::::|::|
1246 rValArgLysSerSerAsnLysAspPhe...T 1256
      ::::::::::::::::::::|::|
3879 AAGCAAGAGAGTCTCCAGCCAGCAGCTGTCATCCCGCTCAGCCACCTA 3928
      ::::::::::::::::::::|::|
1286 hrProGluArgAspLysLysValAspTyrAspSerArgAspTyrSerSer 1272
      ::::::::::::::::::::|::|
3929 CTACAGGACCCCAAGAAA... 3948
      ::::::::::::::::::::|::|
1273 SerLysArgArgAspGluArgGlyGluLeuAlaArgArgLysAspSerPr 1289
      ::::::::::::::::::::|::|
3949 .....GAAGTTCACAAACACACTCCAGTGA 3974
      ::::::::::::::::::::|::|
1289 oProArgGlyLysGluSerLeuSerGlyGlnLysSerLysLeuArgGluG 1306
      ::::::::::::::::::::|::|

```

```

3975 GCCCAAGAAAAAGCAGCCT... 3993
      ::::::::::::::::::::|::|
1306 LuArgAspLeuProLysLysGlyAlaGluSerLysLysSerAsnSerSer 1322
      ::::::::::::::::::::|::|
3994 ....CCACCACAGAAATCAGTCCAGAGCAGCAAGCAAGAAAAAGTG 4038
      ::::::::::::::::::::|::|
1323 ProProArgAspLysLysProHisAspHisLysAlaProTyrGluThrLy 1339
      ::::::::::::::::::::|::|
4039 GCTCCCGCCGCAAGTATCCCTGTAAACAAAA...CCAAAAGAAAAAGA 4085
      ::::::::::::::::::::|::|
1339 sArgProCysGluGluThrLysProValAspLys... 1350
      ::::::::::::::::::::|::|
4086 AAAACCA...CCTCCGCTCAATTAAGCAGAGCAAGAAATCCAGCA 4123
      ::::::::::::::::::::|::|
1351 .....AsnSerGlyLysGluArg 1356
      ::::::::::::::::::::|::|
4124 CTTTAACATCTCTCAGCACTCTCTCCAAATGCAATAGTTCTTAAGCAAAA 4173
      ::::::::::::::::::::|::|
1357 .....GluLysHisAlaAl 1361
      ::::::::::::::::::::|::|
4174 ATTCAGCAGATGAGTCCACAGATCAGATGAGTGAAGTGAAGAGATTTG 4223
      ::::::::::::::::::::|::|
1361 aGluAlaArgAsnGlyLysGluSerSerGlyAlaAsn... 1373
      ::::::::::::::::::::|::|
4224 TGAACGAGAAATGTGTGGAGATGGAGGCTTAGAATCTTGACTCTG 4273
      ::::::::::::::::::::|::|
1373 ..... 1373
      ::::::::::::::::::::|::|
4274 TTCCTATTAACCCAGCGGTGTTCTTCTCTGTCGCCAGTAGTGCGCAT 4323
      ::::::::::::::::::::|::|
1374 .....CysHisVal...Tyr 1377
      ::::::::::::::::::::|::|
4324 GTAGAGTTTGTATTTGCCAAGTGTGTGTGAGCCCTTCCACAAATTTTG 4373
      ::::::::::::::::::::|::|
1377 rLeuThrArgGlnThrLeuPro.TPrArgArgSerTrpLeuLeuLys 1392
      ::::::::::::::::::::|::|
4374 TTTAAGAGAGAACGAGCCGCTCTGTGAGACCAAGCTGGAATTTGGT 4420
      ::::::::::::::::::::|::|

```

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq.us-08-764-233A-1

seq_documentation_block:

Sequence 1, Application US/08764233A
Patent No. 5716849

GENERAL INFORMATION:

APPLICANT: Ligon, James M.

APPLICANT: Schupp, Thomas

APPLICANT: Hill, Dwight S.

APPLICANT: Neft, Snezana

APPLICANT: Ryals, John A.

TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ciba-Geigy Corporation

STREET: 520 White Plains Road, P.O. Box 2005

CITY: Tarrytown

STATE: NY

COUNTRY: USA

ZIP: 10591

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/764,233A

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/729,214

FILING DATE: 09-OCT-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: 1506/CIP6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 49377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Sorangium cellulosum
IMMEDIATE SOURCE:
CLONE: p98/1, pJL3, and pVKM15
FEATURE:
NAME/KEY: misc_feature
LOCATION: 383..760
OTHER INFORMATION: /product= "Sorr"
OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous to type I PKSs such as eryA from Saccharopolyspora erythraea."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 927..19874
OTHER INFORMATION: /product= "Sora"
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKSs that are known to be involved in the synthesis of polyketide compounds."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 942..7115
OTHER INFORMATION: /product= "Module 1 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7203..12884
OTHER INFORMATION: /product= "Module 2 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 13455..19616
OTHER INFORMATION: /product= "Module 3 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19871..46318
OTHER INFORMATION: /product= "SorB"
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKS genes"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19870..24556
OTHER INFORMATION: /product= "Module 1 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 24638..30820
OTHER INFORMATION: /product= "Module 2 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 30881..35446
OTHER INFORMATION: /product= "Module 3 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 35528..40114
OTHER INFORMATION: /product= "Module 4 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 40190..46318
OTHER INFORMATION: /product= "Module 5 of SorB"
FEATURE:
NAME/KEY: misc_feature

```

; LOCATION: 46851..47891
; OTHER INFORMATION: /product= "sorm"
; OTHER INFORMATION: /note= "The protein encoded by the sorm gene is highly
; OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
; OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the
; OTHER INFORMATION: polyketide rapamycin."
US-08-764-233A-1

alignment_scores:
    Quality: 337.00      Length: 1575
    Ratio: 0.499        Gaps: 82
    Percent Similarity: 42.921    Percent Identity: 21.841

alignment_block:
US-09-811-045A-1 x US-08-764-233A-1 ..

Align seg 1/1 to: US-08-764-233A-1 from: 1 to: 49377

26 ASPALAGLUALATYRALALEGLYLYS.....GLUYSPROPR 39
||||| ||| ||||| ||| |||||
7113 GATCGCTCGCCGAGCGCTTCGGACAGAGGCTCTCCGCCAGCAGCAGCG 7122
39 ophleuProGLuGLuProSerSerSerSerGLuGLuAspAspProle. 55
||||| ||||| ||||| ||||| |||||
7163 CGCGGCGCTCCGCGCTCTGCGCCACACAGCAGACAGCAGCGCATGTG 7212
56 PROALAGLULEULEUCYSLEULIECYSLYSASPLIEHETIRAPALAYA 72
||||| ||||| ||||| ||||| |||||
7213 CATTCGTGGCATAGCCCTCGCTCCGCGGCGGCATTTGGCGATGTGCAC 7262
72 VALIIEPROCYSCYSGLYASNSERSECYASPLIUCYSILEARG... 87
||||| ||| ||||| ||||| |||||
7263 GCCTCTT.....GGATTTCCTCCACCAAGAGACGCGAGGCGTGA 7303
88 .....ThrThrLeuLeu 91
7304 GCCCATCCCACTACCCGATGGATGCGGTCGCCCTGACGACCTGATC 7353
92 GLUSERASPLYSHSTHRYSPROTHR.....CYSHISGLINAS 104
||||| ||||| ||||| ||||| |||||
7354 CGCAGCGCAAGGCCAAGAGCTACGCTCGGCATGCTGCCATGCTGACACG 7403
104 NASPVALSERPRO.....ASPALALEULIEALASNLYPHELEUARG 119
||||| ||||| ||||| ||||| |||||
7404 GTGCACTCTTCGACCCCTGATCTTTG...GCATTCACCCCTCGTAGAGC 7450
119 IALIALVALASNAASNPHELYSASNLUHRGLYTYRTHRYLSARGLEUARG 135
||||| ||||| ||||| ||||| |||||
7451 CA.....AACACATGACCCCGCAGCACCGGCTACTGCTCGAAGCTGCT 7494
136 LYSGLIN.....LEUPROPHLEUPEHELEUVA 145
||||| ||||| ||||| ||||| |||||
7495 GGCAGGCGCTCGAAGAGCGGATGCTGCCCTCCACCTCAAGAGATTCT 7544
145 LPROPRO..... 147
7545 CCCACCGCGCTGTCGTCGGCANCRCGCGCCAGCAGTAGACGCCGCGGA 7594
148 .....PROARG 149
7595 ACCGGGCGCGAGGATTCCGAAGCTTACATGTCGAAGCACACTTACGGCT 7644
150 PROLEUSERLINARGASNLLEUINPROARGSERARGSPROILEUENR 166
||||| ||||| ||||| ||||| |||||
7645 CATTTCGCGCGGCGGCTTGCGCTTCACGCTCGGCGTCGAAGGCGCACG 7694
166 GAGINGLINASPROVALVALPHEARGTYRTHRYVALSERPROTHRYSERA 183
||||| ||||| ||||| ||||| |||||
7695 CTTTCGCTCACACCGCTTCTCTCTCTGCTCGCTCGCTCACACCTGCG 7744
183 SPHRYLS.....THRALAGLYSERCYSSERASPSERLY 194
||||| ||||| ||||| ||||| |||||

```



```

717 nlyshLysLysHisArgLysArgAlaGlnGluLysGlyGluGlu 734
720 .....TCGAGCCTCGAGAGGAGT 9229
734 erGluSerPheLeuAsnProGluLeuGluThrSerArgLysCysArg 750
730 CCGGACGCTTCACAGCCCTACAGATGAGAGAGCTAGCTCGCGG 9273
751 GlySerSerGlyLeuAspGluThrLysThrAspThrLeuPheValLeuP 767
9274 CCTCATATGGGCTCTCTCCACCGCTGCTG.....CTGGCATGAA 9314
767 oSerArgAspAspAlaThrProValArgAspGluPrometAspAlaGlu 784
9315 GACGGGTGTGAGATCGCCGCGAGGCCGGAAGCCCTGGAGAAAGAC 9364
784 erLThrPheLysSerValSerAspLysAspLysArgGluLysAspLys 800
9365 CACAGCGCTGCGGCTGACGACGCTTCACCTCCGACATGAGAGGAA 9414
801 ProLysValLysSerAspLysThrLysArgLysSerAspGlySerAla 817
9415 TGCTCGACGACTTCCCGCGCTGCGCCAGCTCACTACATCCCGCA 9464
817 rAlaLysLysAspAsnValLeuLysProSerLysGlyProGluGluLysV 834
9465 CGCATCCCATCATCTCCAAAGTACCGGGCGCGCCAGG..... 9507
834 aLAspLysAspArgLysSerProArg..... 843
9508 .....ACCAGAGCTGCGCTCGCCGACCTACTGGGTCCGCACG 9546
843 ..... 843
9547 TTCGCCACACGTCGCTTCCTGACGCGCTACGTGCTTCACGCCGA 9596
844 .....SerGluProProl 848
9597 GGGGACGCGTCTTCTCGAGCTCGGCGCTCACGCTGCTCTCGCCCT 9646
848 euLysLysAlaLysGluGluAlaThrLysLysSerValLysProSer 864
9647 TGGCAAGAGCGCTCGACAGAGCAAGGACAGTGGCCATGGCGCTCC 9696
865 ...SerSerSerGlnLysAspGluLysValThrGlyThrProArgLysAl 880
9697 TTCACACCTCCGCAAGGAC.....GCGACGACCGCGAGCGTTC 9737
880 aHisSerLysSerAlaLysAspThrArgArgLysGlnPro..... 894
9738 ACCGCGCGCTGCGCTCTCCAGCTGACGGCTCACACCCGACTGGAG 9787
895 .....ArgThrArgSerLysArgThrVal 903
9788 CGCTTCTTCGCCCCCTCGCTCCAGCAAGGTCCCTCCCAACATAG 9837
904 ProLysThrSerSerGlnLys...SerGlnProValArgThrArgArgP 919
9838 CTTTCAGGCGAGCGCTTCTGCTGATGCTTCCAAAGGCAACGCTCC 9887
919 oArgSer.....LeuArgLysLysAsnTyrLeuIleAla 931
9888 GACGTGCGCTCCGAGCGCTGACCTGACCGATACCCGCTGCGCGC 9937
931 rGluLysAsnGluArgGluLysArgLysSerValAspLysAspPhe 947
9938 CGCGTCCCTCGCGCGACCGCATGCTCTCTCACAGAGAGACTCT 9987
948 GluSerSerSerMetLysLysSerLysValGluGlyThrGluLysVal 964
9988 CACTCTGAGACATCGGTGCTCGCGCATCAGCTCAGCTTCG...GTACA 10034
964 sProSerProLysArgLysMetGluLysValGluLysLysGluArgT 981
10035 CCACTCTTCGCG.....GCACCTGCTTCTCGAGCT 10066
981 hrProGluLysAspLysLysLeuLysSerThrThrProAlaLysLysIle 997
10067 TGCCCTGTTGTCGCGCGGTGCGCTCGGCTCGACACCGGTGAAG 10110
998 LysLeuAsnArgGluThrGlyLysLysLysGlyAsnAlaGluAsnAla 1014
10111 .....AGCTC 10115
1014 rThrThrLysGluProSer...GluLysLeuGluSerThrSerSerLysI 1030
10116 ACCCTGAAGACCCCTCGCTCTCCGCTGAAGGCGCCCTCTCGCC 10164
1030 LeuLysGlnGlu...LysValLysGlyLysAlaLysArgLysValAlaLys 1046
10165 ...AGGTGCTGCGTGGGCTTGGACGACGACGACGAG..... 10201
1046 rGluGlySerSerSerThrLeuValAspTyrThrSerThrSerThrG 1063
10202 .....GCCACTCTCTTCAC 10217
1063 LysLysSerProValArgLysSerGluGluLysThrAspThrLysArg 1079
10218 AGCCGACCCCAAGCGCTCTCTCAGAGACGCCCTTGACTGCCACGCGAG 10267
1079 rValLysThrMetGluGluThrAsnAspAsnThrAlaProAlaG 1096
10268 C.....GGCTGCTGCTCCAGCT. 10286
1096 LuAspValIleIleMetIleGlnValProGlnSerLysThrAspLys 1112
10287 .....ACCCGCTCCCTTCCTTCGATCTCCAC 10313
1113 AspPheGlu.....SerGluGluAspVal...ThrThrG 1125
10314 GACTGCGTCTCTCGGCGCGCCACCAAGTAGACACCAAGCCTCTACG 10363
1125 LnProlGlnSerValGlyLysProSerSerLysLysAsnValThr 1141
10364 AACCTCGAAGAGCGCTGCTGCTTACGCGCTTCAGTTCAGGGCTCC 10413
1142 ThrLysPro.....SerAlaThrAlaLysTyrThrGluLysGluSerG 1156
10414 GCTGCTGCTGAGGCGCGGACGAGCTTTCGGAAGCTCAGCTCCG 10463
1156 uGlnProGluLys..... 1160
10464 GACGCGCCCAAAAGATGCCGCTTGCCCTCACCCCGCCCTGCT 10513
1161 .....LeuGlnLysLeuProLysGluLysAsn...HisGluLeu 1172
10514 CGACAGCGCGCTCGACGCGCTTGCCTTTCAGACAGAGCGGACAGCG 10563
1173 MetGlnHisGluLeuArgSerSerLysGly..... 1182
10564 TCGCGCTGCGCTTCTCTGCGGCGAGTCTCTGCGCGCTGCGTGC 10613
1183 .....SerLysSerGluLysGlyArgAlaLysAsp 1194
10614 ACCAGCGTGGCGCTGCGCTTCCACCGCTCGAAGGCAAGCCCGGCTC 10663
1194 rGluHisSerGlySerGluLysAspAsnProAspLysArg..... 1207
10664 GCTGCTCTGCGCGAGCGCGAGCGGACCATGCTCGGTCAAGCGC 10713
1208 ...LysSerLysAlaGlnProAspLysGluSerThrValAspArgLeu 1223
10714 TCGCCAGCGCATACGTCGCGAGAGGCTCCGACCCGAGAGCTTCC 10763
1223 rGluGlnGlyHisPheLysThrLeuSerGlnSerSerLysGluThrArg 1239

```

```

10764 C.....ACCACGATGCCCTCTCCGCGTCGACTGAGCGCA 10788
1240 ThrSerGluLysHisGluSerValArgLysSerAsnLysAspPheTh 1256
      :::::||||| ::::: ::::: :::::
10799 GGTGGCGAGCCCACTACCTCAGCTGAGACCCCAAGCGCCT..... 10840
1236 rProGluArgAspLysLysValAspTyrAspSerArgAspTyrSerS 1273
      ||:::||||| ::::: |||
10841 .CCTTCGTGGCATCGCGCGCTCGACCTCGCCCGAGGTGCTCTCGCC 10889
1273 eLysArgArg.....AspGluArgGlyLulLualArgArg 1285
      :::::||||| ::::: ||||| ::::: |||||
10890 CGCGTGGCGACCTCGCTCCCTCCAGCGCGCTCGA...CCAAGCGCC 10936
1286 LysAspSerProArgGlyLysGluSerLysSerGlyGlnLysSerLy 1302
      ||| ||||| ::::: ::::: ::::: :::::
10937 TTCGGCTCAGCGCTCGTCGTGCTCCCTCATGTGCTAGAACCGCGAG 10986
1302 sLeuArgGluGluArgAspLeuProLysLysGly.....AlaG 1315
      : :::: ||||| ||| ::::: |||
10987 ACCATCATCAGAGCGCCCACTCATCAGCGCGCGCGCTCGCTCGCTG 11036
1315 LuSerLysLysSerAsnSerPro.....ProArgAspLys 1327
      ||| ::::: ::::: |||||
11037 CAAGCGTGGCTGGCGAGAGACGCGCTCGCTCGCTCGCTCGCTGCT 11086
1328 LysProHisAspHisLysAlaProTyrGluThrLysArgProGlyGlu 1344
      ||| ||:::||||| ::::: |||
11087 CACCGGAGCGCGCATCGCTGCGCGCGCGATGAGACGTCAAGGACCTCG 11136
1344 uThrLysPro 1347
      :::::|||||
11137 CTCACGCCCC 11146

seq_name: /cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-144-085-3

seq_documentation_block:
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafson, Claes
; APPLICANT: Betlach, Mary C.
; APPLICANT: Ashley, Gary
; APPLICANT: Julien, Bryan
; APPLICANT: Ziemann, Rainier
; TITLE OF INVENTION: SORANGLUM POLYKETIDE SYNTHASES AND ENCODING DNA
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 30062-20020.20
; CURRENT APPLICATION NUMBER: US/09/144,085
; EARLIER FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010,809
; EARLIER FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 33529
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-144-085-3

alignment_scores:
Quality: 329.00 Length: 1717
Ratio: 0.452 Gaps: 76
Percent Similarity: 42.400 Percent Identity: 19.802

alignment_block:
US-09-811-045A-1 x US-09-144-085-3/rev ..
Align seg 1/1 to reverse of: US-09-144-085-3 from: 1 to: 33529
111 LeuIleAlaAsnLysPheLeuArgGlnAlaValAsnAsnPheLysAsnG 127
||| ||:::||||| ::::: |||

```

```

8324 CTCGCTGCTACACCGCTCCCTGCGTCACACCGAGCGG...CGCATCGA 8278
127 uThrGlyTyrThrLysArgLeuArgLysGlnLeuProProPheLeuPhe 144
      | :::: ||:::||||| ::::: |||
8277 GCCAGCGCTACACCGAGCGCTCTCGCCACAGCG...CAGCATCTCGTAC 8234
144 euValProProArgProLeuSerGlnArgAsnLeuGlnProArgSer 160
      ||| ||::: ||||| ::::: |||
8233 ACTCGCGAGCGCTCGCGCTCTCTCCAGCGGACCTGTGCTCCAGCGCG 8184
161 ArgSerProIleLeuArgGlnAlaAspProValValPheArgTyrThVa 177
      ||| ::::: ||||| ::::: |||
8183 CTCACACCTCTGACCGACCGCACGCTCGGCC...TCCACCGCTCTACACT 8137
177 LysProThrCysSerAspThrLysThrAlaGlySerCysSerAspSerG 194
      ::::: ::::: ||||| ::::: |||||
8136 CGCGCGCAGCTCGCTCGCTCTGCGTGGCGCGCTCG... 8100
194 lYThrLeuSerArgLeuProAlaProSerIleSerSerLeuThrSerAsn 210
      ::::: |||||
8099 .ATGATCACCTCTGACCGGACGCTCGAGGCGCATCACTACCC... 8058
211 GlnSerSerLeuAlaProProValSerGlyAsnProSerSerAlaProAl 227
      ::::: ||||| ::::: |||||
8057 ...GCCGCGGTACGCTCTCGCGCACGCGAGATGACACGAGACCATCT 8011
227 aProVal..... 229
8010 GCCCTCTACGCTCGGCTGCGCTCGCGACGCGTGCCTTCCTTCCACG 7961
230 .....ProAspIleThrAlaThrValSerIle..... 238
7960 COTCTGTACACACGACGACGACGCTGCGCTGCTGCTCTCTCTCGG 7911
239 .....SerValHisSerGluLysSe 245
7910 CGCCACGCGCGAGGTACGCGACGCGCGCGAGCTCTCGCGCACGCTG 7861
245 rAspGlyProPheArgAspSerAspAsnLysLeuLeuProAlaIleAla 262
      ||:::||||| ::::: |||||
7860 GTCAAGC.....ACCTCAGCAGATCCGCAAGCGCTTCGCTTCT 7820
262 euthSerGluHisSer.....LysGlyAlaSerSerIleAlaIleThr 276
      ||:::||||| ::::: |||||
7819 GCACGCGCTCCACACGACGACGCTCGCGCACGAGACATGCGCTGCTCC 7770
277 AlaLeuMetGluGluLysGlyValProGlyThrSerPro..... 289
7769 GCGCTC.....GTCTTGGGCGCTCCACCGAGTACCGCTG 7735
290 ....TyrAsnSerIlePheValGlyGlnSerLeu..... 299
7734 CGGCTGGAACGCG...TACGTGCGAGCTCCACGAGGCGCGCGCACGCG 7688
300 .....LeuHisGlyGlnLeuIleProThrThrGlyProVal 311
7687 CGAGACGACCTGGGCGGACGTCACCTGTGCGCTCGACGCGACGCTGTC 7638
312 ArgIleAsn..... 314
7637 GCCACCGCGCTCGCAGCGCGCGCTCTCTCTGCTCTCGCGCAGGCT 7588
315 .....AlaAlaArgProGlyGlyArgProGly...TyrG 326
7587 CGTCACAAAGCTCGCTGCGCTCTCTCGGAGGACCGCGCGCGCACGCG 7538
326 LuHisSerAsnLysLeuGly.....TyrLeuValSerPro... 337
      ||||| ::::: |||
7537 CGCAGACGACGACCATCGGCGCCACACTGACGCTATGTCCTCACCCCGCC 7488
338 .....ProGlnGlnIleArgArgGlyGluArgSerCy 348
7487 GCCGCAAGCGGTGCGCATCCGCTCCAGGAGCGACCGCTCTCGCACCTG 7438

```

348 slyr.....ArgserLleasnArgly..ArgHisHisSerGlu 360
7437 CCGTACCACCAAGTACTGGCCGCAATGAGGGCTTTGGCCACCGAGCTGCG 7388
361 ArgSerGlnArgThrGlnSerProSerLeuProAlaThrProCysPheVa 377
7387 CCGGTACAGCGCGCTCCACACCGCAGCGGTGGCGGGGTACTGTC..... 7343
377 lProValProProProProLeuTyTProProProHis..... 390
7342 ...ACCCGCGGCGTCACTTCCCGAATCTTCAGACATCCGTCATGTG 7297
391ThLeuPro 393
7296 CCGCGTGTGGAAAGCGGTGCACACACGACAGACGCCGCTGGCGCGCTT 7247
394 LeuProProGlyValProProProGlnPheSerProGlnPheProSerSe 410
7246 GCGCGTTCAGCGCGTGGCCACACCGGAGACCGCGCGTTCGTCGCCGCTC 7197
410 F..... 410
7196 AACACCGTCTGCATGGCGCGTTAGCCCGCATGCTCACTGCCCTTG 7147
411 ...GlnProProThnAlaGlyTyrSerValProProProGlyPhePro 425
7146 GCGCCCGACGCTCGACAGCGCGCGGTGCATACATCCGGCTCGAGGCGTCCA 7097
426 ProAlaProAla..Asn..... 430
7096 CCGACACGATCGCTCCCCCGCGCTGCACCCCTGCATCAGCCGACGGCG 7047
431Ile 431
7046 GCGCACACATCTGGCTGGCGCCGACAGGCTGCACACACCGCCACAGTG 6997
432 SerThrAlaCysPheSerProGlyValProThrAlaHisSerAsnThrMe 448
6996 CCGACGCGTCACTCTCTATCGATGCGCCMAAGACGCGACGGGGTTCA 6947
448 lProThrThrGlnAlaProLeuLeuSerArgGlnGlnPheTyrArgGlu 465
6946 GCCC...CMACACGTCCCACTGACGCTACAGGCCACTTCAAGCGCAAC 6900
465 lAsnAspGlyGlyArgGlnSerLysPheProTyTyrSerGlySerSerTyT 481
6899 AG.....CCGCGCTGGGTGATCTCGCTCGCTCCACCAAGCGCTCTTC 6856
482 SerArgSerSerTyTThrAspSerSerGlnGlyLeuAlaGlnHisLent 498
6855 CMC.....GCTCGCGGCGCGCCGCAACACCACTCTC 6824
498 salLeuThrLeuSer.....ProSerAlaLalaHisThrLeuAspL 512
6823 TCACCCACAGGCTCGAGATGGCGGTCCAGCGCTCGACACCTCGTGCAC 6774
512 eulLeuHisAspHis.....ProHisPro... 519
6773 GCGCAGCGAACAACAGGGGATCACTCTGTAAAGCCCTTCCCATCCGAG 6724
520ProGlnGlnAlaGlnAlaArgSerAlaMetLleVa 531
6723 CCGGTGGCTGCCCTGCCGTGAACAGACGCGAAGCTTCCGCGCTCGCT 6674
531 lHisMetProAspLeuMetAspLleAlaHisAlaArgSerArgSerPro 548
6673 TGGCGGTGCC.....GCTACACACCGCGCATCGGCGCG 6639
548 roTyArgArgArg.....TyArgSerArgSerArgSerPro 559
6638 CCGACAGAGAGCGCGAGACCTTCCACAGTCCCGCGAGCGTCCGCGCG 6589

[illegible]

```

5775  |||  ::::  |||
5775  GCGTGGCCAGAGAGCTTTCAGAGCGCGCTCTCGGCGCATCCCTTGAG  5726
786  .....  ThrpheLysSerValSerAspLysAspLysArg  796
5725  CCGGCTGAACTCCAGCAACAGCGGGGGTCTATCATCCGTCAACCCCG  5676
797  GILuLysAspLysProLysValLysSerAspLysThrLysArgLysSerAs  813
5675  CGGCGAAGCGGCGAGTCCATTCACTGACGCAACGCGCTGCAGCGCAGG  5626
813  pGlySerAlaThrAlaLysLysAspAsnValLeuLysProSerLys...  828
5625  TGCAGCGACACAGCGACGACGACGCGCTGCCACCGCATCGCTGG  5576
829  .....  GlyProGlnGluLysValAspLysAspArgLysSerPro  842
5575  GCCATGCAAGCCCGAGCAGCATAGGCGCACGCGCTGAGAGCAGCGCTCCG  5526
843  ArgSerGluProProLeuLysLysAlaLysGluGluAlaThrLysIleAs  859
5525  CGCTCCCGATGCCCCCGGTACCGCTCCAGCATCTCCAGCCCCG  5484
859  pSerValLysProSerSerSerSerGlnLysAspGluLysValThrGlyT  876
5483  .....  CACCCGTGTAAAGCGCCATAGTCGAAACCCGCGGCCAGATA  5441
876  hrProArgLysAlaHisSerLysSerAlaLysAspThrArgGlnSer  892
5440  CACCCCGG...  TGGCGCTCCCGCTCAGCGCGGCGCAGCGCGCGCAC  5394
893  GlnProArg.....  895
5393  GCTCCATAGCTCCCATCCGCTCCAGCAACACCGCGCTGGGAGATCC  5344
896  .....  ThrArgArgSerLysArgThrValP  904
5343  ATGCGCTGCGCTCTCGGGCGATATCCGAAAGTCCGCGTCGAAACAG  5294
904  rLys.....  ThrSerGlnLysSer  911
5293  GTCGATGTCGCCGAGAAATCCACCTCGCGCGCTAGCTCTTGCCCAACG  5244
912  GlnProValArgThrArgArgProArgSerLeuArgLysIleAsnTyrLe  928
5243  CCTCGCGGTGGGGGTCTAGACCGAAA.....  5217
928  uLleAlaArgGluLysAsnGluArgGluLysArgLysSerValAspL  945
5216  ....  GCGTCTCCAGCGCGAGGAGGCCCTCGATCACGTCCTTCCCTC  5171
945  yAspPheGlnUserSerSerMetLysIleSerLys.....  ValGlu  958
5170  GGGCAAGAGAGCGCATAGTCTTCCGGCGTCTCGACGCTCCGCGCAGC  5121
959  GlyThrGluLeuValLysProSerProLysArgLysMetGluLysAspVa  975
5120  GGGCAGCGCATGATAGATGCGCACGCGCTGCTTCCCGCATGAGCA  5072
975  IGluLysLeuGluArgThr.....  Prog  983
5071  .....  CGACCCCGCTCGGGGCTCCCATCAGCTGCGGTCCG  5034
983  LuLysAspLysIleAlaSerSerThrThrProAlaLysIle.....  997
5033  AAAAGCGCAGCTTCAAGAGCACTCCCGATGCGCGCGGTGCGGTAG  4994
998  .....  Ly  998
4983  TCGAAGACCAAGCGTGGCGGACAGAGCTCCGTCGGGCGGAGAGCG  4934
998  sLeuAsnArgGluThrGluLysLysIleGlyAsnAlaGluAsnAlaSerT  1015
|||:::  |||:::  |||:::  |||

```

```

4933  GTTGGCAGCTCCACCGCCATCAGCGAGTGCAGCCGAGGCGCTTGACA  4884
1015  hrThrLysGluProSerGluLysLeu.....  1023
4883  CTTGCTCGGCGCCACAGCGCTCGCTGCGCTGAGCCCGAGCACCAGCGCG  4834
1024  .....  GluSerThrSerLysIleLysGlnGluLysValLysGlu  1037
4833  ACCTCGCGCCCGACACAGCTCATGAGCGCATTCAGCGCGCTCCGCTCCG  4784
1037  yLysAlaLysArgLysValAlaGlySerGluGlySerSerSerThrLeu  1054
4783  CAGCGCGGAGAGCGCGCGCGCGAGGCGGTCTCGCTCGCTCGCG  4734
1054  aLAsp.....  TyrThrSerThrSerThrGly  1063
4733  TGGAGCGCTTGGCCAGACGCTGGGCGCAACAGCCGACAAACAGCGCGCG  4684
1064  GlySerProValArgLysSerGluGluLysThrAspThrLysArgThrVa  1080
4683  AGCTCGCGCTTGGCGCTCCAGCCACAGCTGCGAGCTGCGAGATCCAGGTG  4634
1080  IleLysThrMetGluGluTyrAsnAsnAspAsnThrAlaProAlaGlu  1097
4633  CATAGGACCAAGCGCTCGCTCAGGCGGTGACAGCGCGCTCCAGCAGCG  4585
1097  spValIleIleMetIleGlnValProGlnSerLysTrpAspLysAspAsp  1113
4584  A.....  MNOCCCTC  4574
1114  PheGluSerGluGluGluAspValLysThrThrGlnProIleGlnSerVa  1130
4573  TTCACGAGACATGACACAGACTCGGTAGCAGTGAACCGTGCAGCTCCG  4524
1130  IGlyLysProSerSerIleIleLysAsnValThrThr.....  1142
4523  CGGC.....  CCCAGCTGTGCGCATGCGCACACCGCGCGCGCTGCAG  4480
1143  .....  LysProSerAlaThrAlaLysTyr  1150
4479  ACACCCCAGGAGAGCTGTCGCGCGGAGCGCGCGCGCGCTGC  4430
1151  ThrGluLysGluSerGluGlnProGluLys.....  LeuGlnLys  1163
4429  CGCGAACGCGTCCGAGAACGTAATCCGCCGCGGTAAATGCCCTGGCGCG  4380
1163  sLeuProLysGluAlaSerHisGluLeuMetGlnHisGluLysArgSer  1180
4379  CATGCCCAACGTAACGCGCGCGCGAGCAAGAACAGCAAGCAAGCGCGCA  4331
1180  erLysGlySerAlaSerSerGluLysGluArgAla.....  1191
4330  ATCAGCTCCCGGTGAGTCTGTCAGGTGACAGCGCGCGCTCCACCTTCG  4281
1192  .....  LysAspArgGluHisSerGluLysSerGluLysAspAs  1203
4280  GCGGCACACCCAGCGAGCGCGCTCGCGCGCTGCTGAATCAACGCCA  4231
1203  nProAspLysArgLysSerGluAlaGlnProAspLysGluSerThrVala  1220
4230  TCATGAGACACAGCCAGGTCGACAGC...  CGCTCAGCGGCGCGCG  4184
1220  spArgLeuSerGluGlnGlyHisPheLysThrLeuSerGlnSerSerLys  1236
4183  CGGTGATGCGCGCGCAGACGCGCGGATCTCTCCGCTTCAGACAGT  4134
1237  GluThr.....  ArgThrSerGluLysHisGlu  1245
4133  CGCAGCGCACCATGACAGACAGTCTGGCGCGGAGCTCGAGCATTTGCA  4084
1245  userValArgLysSerSerAsnLysAspPheThrProGlyArgAspLysL  1262
4083  CAACTGCGGCGCGCGCGCTCCAGCGCGCGCGCGCGAGCTGACAGCA  4034

```



```

1262 yValaAspIyAspSerArgAspIySerSerSerLysArgAspGlu 1278
1279 Arg.....GlyGluLeuAlaArgArgLys.....AspSerProProa 1291
1280 .....GlyGluLeuAlaArgArgLys.....AspSerProProa 1291
3983 CAGACCGCCGACGCGCTCCGTCACACGACGAGTCCGCGAGGTCACAGCC 3934
3984 AGTGCGCTACACCCCGCGCGCGCGCGACAGATGCGCTGTCACCGCCCTGCC 3984
1279 Arg.....GlyGluLeuAlaArgArgLys.....AspSerProProa 1291
1280 .....GlyGluLeuAlaArgArgLys.....AspSerProProa 1291
3983 CAGACCGCCGACGCGCTCCGTCACACGACGAGTCCGCGAGGTCACAGCC 3934
3984 AGTGCGCTACACCCCGCGCGCGCGCGACAGATGCGCTGTCACCGCCCTGCC 3984
1291 rGgLYLysGluSerLysSerGlyGln...LysSerLysLeuArgGluGln 1306
3883 GGGGCGACGACGCGCGCGCGCGCGCGCGACGACGCTCCGCGCGCCCT 3834
1307 ArgAspLeuProLysLysGlyAlaGluSerLysSerAsnSerSerPr 1323
3833 CCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3806
1323 oProArgAspLysLysProHisAspHisLysAlaProTyrGluThrLys 1339
3805 TCACGCGGCTCGGTCGCCACATCGATCAAGCGCGCGCGCGCGCGCG 3756
1340 .....ArgProCysGluGluThrLysProValAspLysAsnSerGly 1353
3755 CTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3706
1354 LysGluArgGluLys.....HisAlaAla 1361
3705 CCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3656
1361 aGluAlaArgAsnGlyLysGluSer.....SerGlyA 1372
3655 ACCAGCTCGACAGCGTCGAGCGCGCGCGCGCGCGCGCGCGCGCG 3606
1372 laAsnCySHisValTyrLeuThrArgGlnThrLeuProThrArgAsp 1388
3605 CGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3556
1389 TrpLeuLeuGlyArgTrpArgAlaProSerSerArgAsnPro 1403
3555 GCGGTC.....AGCTGCGACGCGT.....TCGTGCGATCAACA 3523
seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:us-08-258-261B-6
seq_documentation_block:
: Sequence 6, Application US/08258261B
: Patent No. 5639949
: GENERAL INFORMATION:
: APPLICANT: Schnupp, Thomas
: APPLICANT: Ligon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,261B
FILING DATE: 08-JUN-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-258-261B-6

alignment_scores:
Quality: 323.50 Length: 1578
Ratio: 0.495 Gaps: 83
Percent Similarity: 41.445 Percent Identity: 22.687

alignment_block:
US-09-811-045A-1 x US-08-258-261B-6 ..
Align seg 1/1 to: US-08-258-261B-6 from: 1 to: 28958

118 ArgGlnAlaValAsnAspPheLysAsnGluThrGlyTyrThrLysArgLe 134
7748 CGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7797
134 uArgLysGlnLeuProPheLeuPheLeuVal.....ProProp 148
7798 GCCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7847
148 roArgProLeuSerGln...ArgAsnLeuGlnProArgSer..... 160
7848 CGTGGCGGACATCGACTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 7897
161 .....ArgSerProLeuLeuAr 166
7898 GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7947
166 g.....GlnGlnAspProValAlaPheArgTyrThrValSerProT 180
7948 ACCAGCGAAGCGCGCGCTCCGCGACATGGCCGCGCGCGCGCGCG 7997
180 hCysSerAspThrLysThrAlaGlySer.....CysSerAspSer 193
7998 CAAGCTCGGAGCGCGCTCGGAGACGACGATCGTGTGCTGTGCG 8047
194 GlyThrLeuSerArgLeuProAlaProSerLysSerLeuThrSerAs 210
8048 TGGTGGCGCGACGAGCGCGC.....TCGTC 8073
210 nGlnSerSerLeuAlaProProValSerGlyAsnProSerSerAlaPro 227
8074 CTCGCGCAGAGCGCGCTCCCATGTG..... 8101
227 laProValProAspIleThrAlaThrValSerLysSerValHisSerGlu 243
8102 .....ACCCGACAAAGGCGTCTCGATCTCGGTCTGATTCGATATG 8146

```


9683 CCGTCTCC.....TGTCGGCCAGAA..... 9703
760 rAspThrLeuPheValLeuProSerArgAspAspAlaThrProValArga 777
9704GCAGAGCCGCGTCGCCGCCAGCGCAA 9731
777 spGLuProMetArgAspAlaGluSerIleThrPheLysSerValSerAspLys 793
9732 GCGGCTCCGCCAGCACACTCTCTGCCCCACAGACACTCGCCCTTATCGARG 9781
794 AsPLysArgGLuLysAspLysProLysValLysSerAspLysThrLysAr 810
9782 TGCGCTATTGCGAGGCCAACACCCGGGCCCACTTGAGAGACCGCCCGCT 9831
810 gLysSerAspGLysSerAlaThrAlaLysLysAspAsnValLeuLysProS 827
9832 CTCCTGGCCCGGAGCCGGAGC.....A 9854
827 eLysGLYProGLInGLysValAspGLYAspArgGLuLysSerProArg 843
9855 GCTCTCTTCGCGCTCG.....ACTGCGTCGCCACGG 9886
844 SerGLuProProLeuLys.....LysAlaLysGLuGLuAlaThrLys11 858
9887 ACAACGCCGCCCCGAGCAACCGTTCTTCGGCGCGGAGGAACACAG..... 9931
858 eaSPserValLysProSerSerSer.....SerGLInL 869
9932SCAAGTCGTCTCTCTTCTTCTGGGCAAGCTCGCACTGGGA 9974
869 ys.ASP.....GLuLysValThrGLYThrProArgLysAl 880
9975 AGGATGGCCCTCCCTCCCTGCTGACTCTCTCGCGGGTCTTCCGCGCTCAC 10024
880 aHisSerLysSerAlaLysAspThrArgArgGLInSerGLInProArgThra 897
10025 TCGAAGCATCGACGAG.....CGCGCTCGCTCTCACCTGACATGG 10065
897 rGArgSerLysArgThrValProLysThrSerSerGLInLysSerGLInPro 913
10066 AGCCTGCTTCGCGCTCTGCGCGCGCACGAGGCGCCCTCCCTCGACCG 10115
914 ValArgThrArgArgProArgSerLeuArgLysIleAsnTYr..... 927
10116 CPTCGACGCTGACAGCCCGCCTTGGCCGTGATGGCTCTCTGCGCG 10165
928 ...LeuIleAlaArgGLuLysAsnGLuArgGLuLysArgLysSerV 943
10166 CCTCTTGCGCTGCTCGCGCGTGCAGCCCGCGCGCTGCGGCACACAC 10215
943 aLAspLysAspPheGLu...SerSerSerMetLysIleSerLysValGLug 959
10216 CAGGCGCAGATGCGCGCGCGCTGTCTGCAAGCGCTCT..... 10255
959 LYThnGLuIleValLysProSerProLysArgLysMetGLuGLYAspVal 975
10256CCCTCGAGACGCGCGCGCGCATCGCCGCCCTTCG 10288
976 GLuLysLeuGLuArgThrProGLInLysAspLysIleAla.....Se 989
10289 GCAGAGAAACCGCTCACACCGTTCGCGCGCGCAACGCGGCAATGCGCCGCTC 10338
989 rSerThrThrPro.....AlaLysLysIleLysLeuAsnArgGLuThnrg 1004
10339 GAGCTCGCGCGCTCGACCTCGACACTACACTACGCTCTCGTGGGCGACAG 10388
1004 LYLys..... 1005
10389 GCTTTCACACGCCCGCGCTCACAGCCCCAGGGCTACCTCGATCCGCGG 10438
1006Lys11 1007

10439	AGCCGCGCCGCTCGACGGCGTCTGACGTCTTACCGCCACCAAGGtG	10488
10493	AGClysnAlaglAsnAlaSerThrLySgluProSerGluLySleuG	1024
1007	ecLySnAlaglAsnAlaSerThrLySgluProSerGluLySleuG	1024
10489	TTGCGCCGCAAGATCGCGCTCGACTACG.....CTCCGACTCCGCCCA	105332
1024	luserTherSerGluLySleuSgIngluLySValLySglLySnAlaLyS	1040
10533	GAATGCAGCCGCTCGAAGACGACTGCGCCGACGtC.....	10567
1041	ArgLySValAlaglYSerGluLySerSerThrLeuValAsPtyrTh	1057
10568TAGCCACATGCTCTCGACGtGC	105933
1057	rSerThrSerThrInglYSerProValArgLySerGluLyST	1074
10594	GAGCTCCCTTATTATTCGACCTCCACCGCACAGGCTGCGAGCTCGCA	106433
1074	hAspThrLySArgThrValIlleLySthMetGluGluLytrAsnAsp	1090
10644	GCTCGACGGCGGCTACTGTTATTCGAACCTCCGCGAAGCCGCTCTTCT	106533
1091	AsnThrAlaProAlaGluAspValIlleIleMetIleGluValProGluSe	1107
10694	CGAGCGCACCGACGGCGCTCTCGACGAATGGCATCGCTTCCG.....	10738
1107	rlYstrpAspLySAspAspPheGluSerGluGluGluAspValYstrHt	1124
10739TCGAGGT	10745
1124	hrcAlaProIle.....GlnSerValGlyLySProSerSerIle	1136
10746	CAGCCCCATCCCGTCTGTCAGCGCTGCGCTCGCGACAGCTCGAGCGGT	10795
1137	lIeLySAsnValThrThrLySProSerAlaThrAlaLySTyrThGluLy	1153
10796	CACCGCTCGATC.....CCGTCGTCTGCGCTCCATTGTGACAGAGA	10836
1153	sgLuserGluGlnProGluLySleuGlnLySleuProLySgluAlaSerH	1170
10837	GAAAGCCACACTGGCC.....	10852
1170	IsGluLeuMetGlnHisGluLeuArgSerSerLySglYSerAlaSerSer	1186
10853GCCTGCTCTCTCTGCGCGGAGCTCTTACCCGAGCGCTCGCGC	10897
1187	GluLySglYArgAlaLySAspArg.....	1194
10898	TCGACTGGAAGACyTTCTTGCGGCCCTTACGTCGCCCGCAAGGTCTCCCTC	10947
1195GlnHisSerGlySerGluLySAspAsnPro....	1204
10948	CCCACTTACCCTTCACGAGGAGGCGTTCTGCTCGACGTCTCCACGGA	10997
1205AspLySArg	1207
10998	CGAAGCTTCCGACGTCGCTCCGACGGCTTACCTCGCGCCACCAATCC	11047
1208	LySerGlyAlaGlnProAspLySgluserThrValAspArgLeuSerCl	1224
11048	CGCTGCTCGGCGCCGCGCTTCGCTTCGCCGACCGCGGtGCGCTTCTCTT	11097
1224	uGlnGlnHisPheLySthrLeuSerCluserSerLySgluThrArgTrHs	1241
11098	ACAGGCGCGCTCTCCCTCGCAAGACACCCGTGCTCGAAGCCATGCGCT	11147
1241	erGluLySglYSerValArgGly.....SerSerAsn.....	1252
11148	CTTCCGACACCACTTCATACGGGACCGCGCTTCTTCGAGCTCGCCCTGC	11197
1253LysAspPheThrProGlyArgAspLySValAs	1264
11198	ACGTGCGCCACGCGCGCTCTCGACACCGGTGCGAAGAGTCAAGC.....	11242

```
1264 pTyraSPSerArgSPtyrSerSerSerLysArg..... 1276
      |||||
11243 .....TCGAGGCCCTCTCGCTCTCCATCGCAGGACACCGTCCTCT 11285
1276 ..... 1276
11286 CCAGATCTCCGTCGGGCCGTCGAGACGACGAGGAGGCGCTCTCTT 11335
1277 .....AspGluArg..... 1279
11336 TCCTATAGCCGACAGAGAGCGCCTTCAGATGCCCTGAGCTGCCAC 11385
1280 ...GlyGluLeuAlaArgArgLys.....AspSerProProArgGlyL 1293
      ::::|
11386 GCCAGCGGCTCTCTCGCGCGGAGACCCATCCCTCTCGCGCATCTCCA 11435
1293 ySGlSerLeuSerGlyGlnLysSerLysLeuArgGlu..... 1306
      ::::|
11436 CGAGTGGCTCCCTCGAGTGCCTCCGCTGAGCCTCGAAGCCTCTACG 11485
1307 ArgAspLeuProLysGlyGlyAlaGluSerLysLysSerAsp.....Se 1321
      ::::|
11486 CAACCTCGCCCAACCTCGGCTTGCCTACGCGCCCGAGTTCAGAGGCTC 11535
1321 fSerProProArgAspLysLysProHisAspHisLysAlaProTyrGluT 1338
      |:::|
11536 CGCTCCGCTACAGGCGGCGAGAGCTCTTGGCGAAGCCAGAGTCCC 11585
1338 hLysArgPro.....Cys 1342
      |||||
11586 GGAAGCGCGCGAAAGATGCCCGCGGTTTGCCTCCACCTGCGCTGC 11635
1343 GluGluThrLysProValAspLysAsnSerGlyLysGluArgGluLysH 1359
      ::::|
11636 TCGACAGCGCCCTGCATGCACTGGCCTTGAGAGACGAGCAGAGAGGACG 11685
1359 sAlaAlaGluAlaArgAsnGlyLysGluSerSerGlyAla..... 1372
      ::::|
11686 GTCGCTCTGCTCTCTCGTGAAGGAGTCTGCGCTCGCTCGGTGC 11735
1373 .....AsnCysHisValTyrLeuThrArgGlnThrLeuProTTrp 1385
      |||
11736 CACCACTTGCCTGCGCTTCACCGTCCAAAGGTGAATCTCCCTCT 11785
1386 ArgArgSerTrp.....LeuLeuGlyArgTTrpArgAr 1396
      |||||
11786 CGATGCTCTGCGCGACCGCGAGGTGACCTCTTGCCTCGGTGCAGGCG 11835
1396 gAlaProSerSerArgAsnPro 1403
      |:::|
11836 CTCGCCATGCGGAGCAGCTCCG 11857
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2002, 17:23:55 ; Search time 46.67 Seconds
(without alignments)
2890.710 Million cell updates/sec

Title: US-09-811-045a-1
Perfect score: 7374
Sequence: 1 MMEVDPNNKGMALNTGKTY.....WRRSMLGRMRAPSSRNPS 1404

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6138.5	83.2	1560	T42727	proliferation pote
2	3500.5	47.5	948	A57640	retinoblastoma bin
3	1305.5	17.7	529	T50609	hypothetical prote
4	480.5	6.5	1877	T21861	hypothetical prote
5	383	5.2	3488	T34418	hypothetical prote
6	374.5	5.1	5327	T33564	microtubule-associ
7	363.5	4.9	6642	T29757	protein UNC-89 - C
8	349	4.7	990	T51618	nucleolar phosphop
9	342	4.6	1507	B47328	natural killer cel
10	341	4.6	3968	A44265	trithorax homolog
11	340	4.6	2738	E88320	protein F07A11.6 [
12	339	4.6	2722	T20532	hypothetical prote
13	337.5	4.6	1020	QFHHH	neurofilament trip
14	336	4.6	2526	T20531	hypothetical prote
15	327	4.4	2187	T30826	nascent polypeptid
16	326.5	4.4	1390	S51364	sperm tail-specifi
17	325	4.4	1526	A45605	mature-parasite-in
18	324	4.4	971	T19431	hypothetical prote
19	321	4.4	2464	QRMSP1	microtubule-associ
20	320	4.3	1165	T16420	hypothetical prote
21	320	4.3	1280	T00365	hypothetical prote
22	316.5	4.3	606	A43427	neurofilament trip
23	315	4.3	3869	A48205	All-1 protein +GPE
24	314	4.3	1173	T31421	C-terminal domain-
25	313.5	4.3	2649	T51023	hypothetical prote
26	313	4.2	1274	T16251	hypothetical prote
27	310	4.2	1791	T02345	hypothetical prote
28	304.5	4.1	852	T06310	hypothetical prote
29	302.5	4.1	1200	A46194	neurofilament prot

30	302	4.1	5170	2	T15348	hypothetical prote
31	299.5	4.1	2364	2	A56577	microtubule-associ
32	298.5	4.0	1390	2	T14004	trita protein - sli
33	297	4.0	1192	2	A71623	probable secreted
34	296.5	4.0	3147	2	T18674	hypothetical prote
35	296.5	4.0	3938	2	T42761	Bassoon protein -
36	296	4.0	1320	2	JC5630	trcof1 protein - mo
37	295	4.0	749	2	A45294	Babini ring 2.1
38	294.5	4.0	1641	2	T18614	helicase II - huma
39	294	4.0	699	2	T38073	nucleolar phosphop
40	294	4.0	1403	1	A47328	natural killer cel
41	293.5	4.0	990	2	H66293	protein T24D18.4 [
42	290	3.9	1432	2	B85431	trichohyalin like
43	289.5	3.9	2845	2	T49505	adenomatous polyo
44	288	3.9	1870	2	S37671	MHC class III hist
45	288	3.9	1871	2	D96796	probable heat shoc

ALIGNMENTS

RESULT 1	T42727	proliferation potential-related protein - mouse
C:Species:	Mus musculus (house mouse)	
C:Date:	11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 02-Sep-2000	
C:Accession:	T42727	
R:Witte, M.M.; Scott, R.E.	submitted to the EMBL Data Library, November 1998	
A:Reference number:	Z22246	
A:Accession:	T42727	
A>Status:	preliminary; translated from GB/EMBL/DBJ	
A:Molecule type:	mRNA	
A:Residues:	1-1560 <WTT>	
A:Cross-references:	EMBL:U83913; NID:93859884; PID:93858885; PIDN:AAC72432.1	
A:Experimental source:	strain Balb/C	
C:Genetics:		
A:Gene:	P2P-R	
A:Function:	Involved in hnRNP association and Rb1 binding	
A:Superfamily:	RING finger homology	
F:57-107/Domain:	RING finger homology <RRN>	
Query Match	83.2%; Score 6138.5; DB 2; Length 1560;	
Best Local Similarity	85.4%; Pred. No. 3.1e-230;	
Matches 1207; Conservative 36; Mismatches 133; Indels 37; Gaps 7;		
QY 1	MMEVDPNNKGMALNTGKYATPTDAEAYATGKKKPPFLPEPPSSSEEDDPIPAELL 60	
DB 1	MMEVDPNNKGMALNTGKYATPTDAEAYATGKKKPPFLPEPPSSSEEDDPIPAELL 60	
QY 61	CLICDIMTDAVAVIPCCGNSGSCDECIRTTLLSDKRTCTCHQNDVSPDALIANFLROA 120	
DB 61	CLICDIMTDAVAVIPCCGNSGSCDECIRTTLLSDKRTCTCHQNDVSPDALIANFLGRA 120	
QY 121	VNMFNKGTYTRKRLKQLPFLFLVPPRPRLSQRNLOPSSRSRPIRLQODPVYFRTVST 180	
DB 121	VNMFNKGTYTRKRLKQLPFLFLVPPRPRLSQRNLOPSSRSRPIRLQODPVYFRTVST 180	
QY 181	CSDTAGSCSDSGTSLRPLPAPSSISLTNOSSLAPVSGNSSAPAPVDDTAATVSI 240	
DB 181	CSDTAGSCSDSGTSLRPLPAPSSISLTNOSSLAPVSGNSSAPAPVDDTAATVSI 240	
QY 241	HSEKSDGPPRSDNKLPLPAALITSEHSKGASSIATLAMEEG---VPGTSPMNSIFVG 296	
DB 241	HSEKSDGPPRSDNKLPLPAALITSEHSKGASSIATLAMEEGGYQVPLGTDS---LLG 278	
QY 297	QSLRHQQLIPTGPVIRINARPGGPGGHEHNSKLGIVLSPQQRIRGRGRSYRSTNRR 356	
DB 297	QSLRHQQLIPTGPVIRINARPGGPGGHEHNSKLGIVLSPQQRIRGRGRSYRSTNRR 338	
QY 357	HHSERQRTQSPSLPATPCFVFPVPPPLXPPLPPLLPPLPGVPPQFSPQFSSQPPTAG 416	

```
||||| 339 HHSRSQRTQGPSPALPAPVFPVPPPLPPLPPLPPLPGLVLPQSPQPPQPPAG 398
417 YSVPPGFPAPANISTACFSPGYPTAHSNTMPTQAPLISREEFYRQONKQGESKPPY 476
399 YSVPPGFPAPANISTACFSPGYPTAHSNTMPTQAPLISREEFYRQONKQGESKPPY 458
477 GSGSYSRSSYTD--SQGLAQIHAIHLTSPSAHTLDLHDHPH---PEEAARSAMIV 531
459 SGSSYSRSTYDSSKRSRSGTSRSGTSRSGTSRSGTSRSGTSRSGTSRSGTSRSGTS 518
532 HMPDLMDIAHARSPPRRRRRSRSPPEFRGSPPTKRNYPREKEKREYNNRRREVPP 591
519 H-----GYHSRSRSPPTKRIHHSRSGTQAFRGOSPTKRNYPREKEKREYNNRRREVPP 573
592 YDIKAYGRSVDFRDPPEKERYREMERKRYRMRYKYKGYAVGAQPPRSANREDFSPERL 651
574 YDIKAYGRSVDFRDPPEKERYREMERKRYRMRYKYKGYAVGAQPPRSANREDFSPERL 633
652 LPLNIRNSPPTGRREDYAGQSHRNRLGNYPEKLISTRDSHNAKDNPKSESESNVP 711
634 LPLNIRNSPPTGRREDYAGQSHRNRLGNYPEKLISTRDSHNAKDNPKSESESNVP 693
712 GDGKGNHKKRRKRRNEKEGSESEFLNPELLETSRKCRSGSGIDETDTLFLVLPSCRDD 771
694 GDGKGNHKKRRKRRNEKEGSESEFLNPELLETSRKCRSGSGIDETDTLFLVFPSCRDD 753
772 ATPPRDEPMDEESTIFKSVSDKREKDKPKVKSQDKTKRKGDSATAKDNVLPSPKQ 831
754 ATPPRDEPMDEESTIFKSVSDKREKDKPKVKSQDKTKRKGDSATAKDNVLPSPKQ 813
832 EKVGDREKSRSEBPEPLKAKAEATKIDSVKSSSQDEKVTGT PKRAHSAKADTRRQ 891
814 EKVGDREKSRSEBPEPLKAKAEATKIDSVKSSSQDEKVTGT PKRAHSAKAKEHQA 873
892 SQPTRSKRKYPTTSQKSPVTRRRSLKTNILYIAR-EKEREKRAKSVQKDESS 950
874 KPADEKAKKCSQDISEKSPASDEKAKKPEKKLLDSKEKRRKREEKSVKDESS 933
951 SMKISKVGEIVPSPRRKKEGVEKLEPTPEKDKIASSTPAKKITLNETGKKIGNA 1010
934 SMKISKVGEIVPSPRRKKEGVEKLEPTPEKDKIASSTPAKKITLNETGKKIGNA 993
1011 ENASTTKEPSEKLESTSSKIKOEKVKGKAKRVAGSESSSTLVDTYSTSGSPVAKS 1070
994 ENASTTKEPSEKLESTSSKIKOEKVKGKAKRVAGSESSSTLVDTYSTSGSPVAKS 1053
1071 EKKDTRKTVKTMEEYNNNDTAPAEVYIMIQVPQSKMDKDDFESEBEDVKTQPIQSV 1130
1054 EKKDTRKTVKTMEEYNNNDTAPAEVYIMIHVPQSKMDKDDFESEBEDVKTQPIQSV 1113
1131 GKPSIITKNVTTPSATKATYEKSEDOPEKQKLPKEASHELMJOHELRSKGSASSEGR 1190
1114 GKPSIITKNVTTPSATKATYEKSEDOPEKQKLPKEASHELMJOHELRSKGSASSEGR 1173
1191 AKDBHSGSEKDNPKRRSGAQPDKESTVDRLSEQGHFKTLQSQSKETRTSEKHESVAGS 1250
1174 AKDBHSGSEKDNPKRRSGAQPDKESTVDRLSEQGHFKTLQSQSKETRTSEKHESVAGS 1233
1251 SNKFTTGRDKKVDYDSYSSSKRDERGELARRKDSPPGKESLSQOKKLEERDL 1310
1234 SNKFTTGRDKKVDYDSYSSSKRDERGELARRKDSPPGKESLSQOKKLEERDL 1293
1311 KKGAEKSSNSPPRDKKPHBHKAPYETKRPCEETKPPDKKSGKEREKHAAREKNGESS 1370
1294 KKGAEKSSNSPPRDKKPHBHKAPYETKRPCEETKPPDKKSGKEREKHAAREKNGESS 1353
1371 GANCHVYLTRQTLFWRBSWLLGRWRAPSSSRNP 1403
1354 GGLPCILNPDLPMEKELAVGVEKSAVVKRP 1386
```

```
A57640
retinoblastoma binding protein RBO-1 - human
C:Species: Homo sapiens (man)
C:Date: 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change 17-Mar-2000
C:Accession: A57640
R:Sakai, Y.; Saijo, M.; Coelho, K.; Kishino, T.; Nikiwa, N.; Taya, Y.
Genomics 30, 98-101, 1995
A:Title: cDNA sequence and chromosomal localization of a novel human protein, RBO-1 (
A:Reference number: A57640; MUID:96129310
A:Accession: A57640
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-948 <SAK>
A:Cross-references: GB:A85133; NID:9728590; PIDN:CA59445.1; PID:9755748
C:Genetics:
A:Gene: GDB:RBBP6
A:Cross-references: GDB:626076
A:Map position: 16p12-16p11.2
C:Superfamily: RING finger homology
C:Keywords: tandem repeat; zinc
F:79-129/Domain: RING finger homology <RNG>

Query Match 47.58; Score 3500.5; DB 2; Length 948;
Best Local Similarity 73.28; Pred. No. 1,3e-128;
Matches 691; Conservative 44; Mismatches 130; Indels 79; Gaps 10;

QY 1 MMEYKDPNMKGAMLTNNGKVAIPTDAEAYAGKKEKPPLEPSSSSSEEDDIPAEEL 60
DB 23 MMEYKDPNMKGAMLTNNGKVAIPTDAEAYAGKKEKPPLEPSSSSSEEDDIPAEEL 82
QY 61 CLICDINTDAVVIPICCGNSCDECIPTLTLESDDHTCPTCHQNDVSPDALIANKFLRQA 120
DB 83 CLICDINTDAVVIPICCGNSCDECIPTLTLESDDHTCPTCHQNDVSPDALIANKFLRQA 142
QY 121 VNNFNKNETGYTKRLKQPLFLVPPRPRLSQNLDPKRSNSPILRQODPVYFYTYTSP 180
DB 143 VNNFNKNETGYTKRLKQPLFLVPPRPRLSQNLDPKRSNSPILRQODPVYFYTYTSP 202
QY 181 CSDTKTGTAGSCSDGTLRLPAPSISSLTSSNOSLAPVSGNPSAPAPVDITATVSI 240
DB 203 -----HAPSISLSLSSNOSLAPVSGNPSAPAPVDITATVSI 244
QY 241 HSEKSDPFRSDNKLPAALATSEHSGASSIATLAMEEK-----VPGTSPWNSIFVG 296
DB 245 HSEKSDPFRSDNKLPAALATSEHSGASSIATLAMEEKGYQVYVLTGPS-----LLG 300
QY 297 QSLHGGILPTTGIVRTINARPGGGRGWEHSHNKLGLVSPQOIRRGESCYRSINRGR 356
DB 301 QSLHGGILPTTGIVRTINARPGGGRGWEHSHNKLGLVSPQOIRRGESCYRSINRGR 360
QY 357 HHSRSQRTQGPSPALPAPVFPVPPPLPPLPPLPGLVLPQSPQPPQPPAG 416
DB 361 HHSRSQRTQGPSPALPAPVFPVPPPLPPLPPLPGLVLPQSPQPPQPPAG 420
QY 417 YSVPPGFPAPANISTACFSPGYPTAHSNTMPTQAPLISREEFYREQ----- 465
DB 421 YSVPPGFPAPANISTACFSPGYPTAHSNTMPTQAPLISREEFYREQ----- 480
QY 466 -----NDKRE-----SKFYSGSSYSRSTYD--SQGLAQIHAI 501
DB 481 KLDEFNDFAKELMEYKIKERRRSFSRSPYSGSSYSRSTYSGSSYSRSTYSGSSYS 540
QY 502 LSPSAHTLDLHDHPH-----PEEAARSAMIVHMDLMDIAHARSPPYRRYRSR 557
DB 541 RSPFSRSTYSGSSYSRSTYSGSSYSRSTYSGSSYSRSTYSGSSYSRSTYSGSSYS 595
QY 558 SPPEFRGOSPTKRNVPREKEREYFNRYREVPPYDIKAYGRSVDFRDPPEKERYREME 617
DB 596 SPQAFRGOSPMKRNVPQGETEREYFNRYREVPPYDMKAYGRSVDFRDPPEKERYREME 655
QY 618 RKYREWEKYYKGYAVGAQPPRSANREDFSPERLLPLNIRNSPPTGRREDYAGQSHRN 677
```

Db 656 RRYREWEKYYKGYAAGQAPRPSANRENPSPERFLPLNIRNSPTFGRREDYVGQSHRS 715
QY 678 RNLGNTYPEKISTRDSSNAKDNPKSEKESENVPGDGNKKHKKRRRNEEGESESF 737
Db 716 RNIGSNYPEKLSARDGHQKNTKSEKESENAPDGGKGNKKHKKRRR---KGESEGEF 772
QY 738 LNPLELETSRRCRSGSSGIDETDTDFLPLPSRDATPVPRDEPMDSITFKSVSKDRE 797
Db 773 LNPLELETSRKSREPTGVENKTDLSLFVLPSPRDATPVPRDEPMDSITFKSVSEKDKRE 832
QY 798 KDKPKVSKDKTRKRSKDSGATAKKDNVLPKSPQEKVDGDRKESPRSEPLTKAKEEATK 857
Db 833 RDKPRKAGDKTRKKNKDSGAVSKENIVAKPAQPGQEKVDGDRDLDLNLQLKPKRRRLR 892
QY 858 IDSVKPSSSSOKDEKVTGTPPKAHKSAKADTRKOSQPTRRSKR 901
Db 893 LTLNLHPLRLRRMKKSLEPPEKL-----TLNQOKTPRRKTSQR 930

RESULT 3
T50609
hypothetical protein DKFzp761B2423.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T50609
R:Bioecker, H.; Bioecker, M.; Brandt, P.; Mewes, H.W.; Well, B.; Wiemann, S.
submitted to the Protein Sequence Database, June 2000
A:Reference number: 225143
A:Accession: T50609
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-529 <AAA>
A:Cross-references: EMBL:AL359564
A:Experimental source: adult amygdala; clone DKFzp761B2423
C:Genetics:
A:Note: DKFzp761B2423.1

Query Match 17.7%; Score 1305.5; DB 2; Length 529;
Best Local Similarity 77.9%; Pred. No. 7.5e-44;
Matches 250; Conservative 32; Mismatches 38; Indels 1; Gaps 1;
QY 1052 TLVDYTTSTSTGSPVARKSEKTDTRKYIKTMEYIYNDNTPAEDVIIMIOVPSKMDK 1111
Db 1 TLVDYTTSTSTGSPVARKSEKTDTRKYIKTMEYIYNDNTPAEDVIIMIOVPSKMDK 60
QY 1112 DDFESEEDVKTQPIQGVKPSIIKNVYTKPSATKYTEKESQPEKIQKLPKEASHE 1171
Db 61 DDFESEEDVKTQPIQGVKPSIIKNVYTKPSATKYTEKESQPEKIQKLPKEASHE 120
QY 1172 LMOHLEBSKSGASSEKGRADREHSGSEKNDPKRKSQAOPDKESTYDRLSEOGHFTL 1231
Db 121 LMOHLEBSKSGASSEKGRADREHSGSEKNDPKRKSQAOPDKESTYDRLSEOGHFTL 180
QY 1232 SOSSEKTSRSEKESVNRSSNKFPGRDKKYDVSPOVSSSKRRDERGELARRRDSPPR 1291
Db 131 SOSSEKTSRSEKESVNRSSNKFPGRDKKYDVSPOVSSSKRRDERGELARRRDSPPR 240
QY 1292 GKEISLQGRSKLREERDLPPKKG-AESKSSNSPPRDKRPHDKAKAYETRRPCEETPYDK 1350
Db 241 NKDSASGQKNKPRERDLPKKGTDGSKSSNSPSRDKRPHDKAKAYETRRPCEETPYDK 300
QY 1351 NSGKERKHAALFARNGKSSG 1371
Db 301 NSGKERKHAALFARNGKSSG 321

RESULT 4
T21861
hypothetical protein F36F2.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21861

R:Cottage, A.
submitted to the EMBL Data Library, November 1996
A:Reference number: 219479
A:Accession: T21861
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1877 <WTL>
A:Cross-references: EMBL:Z81532; PIDN:CA804326.1; GSPDB:GN00019; CESP:F36F2.3
A:Experimental source: clone F36F2
C:Genetics:
A:Gene: CESP:F36F2.3
A:Map position: 1
A:introns: 49/3; 86/2; 112/3; 139/2; 235/3; 284/3; 436/3; 507/2; 566/2; 678/2; 1000/3

Query Match 6.5%; Score 480.5; DB 2; Length 1877;
Best Local Similarity 19.0%; Pred. No. 2.2e-11;
Matches 276; Conservative 168; Mismatches 400; Indels 605; Gaps 51;

QY 1 MME--VKDPNMGAULTNKGKAIPTI--DAEAYAIKKKKKPPFLPEERSSSEEDDPIR 56
Db 249 IMETTVDDPD--AAHHBPGKYVIFIMHWKARQETLARKNE-----DGSSPAQSRKVP 300
QY 57 AELLCLICKDITDAVVIIPCCGNSCDECIRTTLESDEKHTCP--TCHONDVSPALIAN 114
Db 301 PELLCPICQSLFEKAIIVISCCGNSYCADCIEARIIDPDNOKPGADCGK-DISTSIIPN 359
QY 115 KFLQAVNNKNEGTYYTKRLKQDLPFLFVPPPPRLSQRLQDPSRSPILRQDPVFR 174
Db 360 KTLRDAAMAAWLSATGPGAPTPPQIVP-----BPEQIR 391
QY 175 YTVSPGSDTKRAGSCSDSGTSLRLPAPSISSLTNSQSLAPVSGNSPAPVPDITA 234
Db 392 IRIG-----LKAPS-----SSQSIITP----- 408
QY 235 TVSISVHSEKSDGPPRDSNDKLLPAAALTESEKSGASSIATALMEKKGPGTSPWNSIF 294
Db 409 -----SGISPGSTIV 418
QY 295 VGOSLHGGILPTGPVRINARPGGGRPEWHSNKLGYLVSPQOIRRGERSCYRSINR 354
Db 419 OQOFTL-----TSVSS 429
QY 355 GRHHSEPRGORTSPSLPAPPCFPVP--PPLYPPTPTLPLPGVPP--PQSPQFPSS 410
Db 430 GTSLSAQPSNV--NPSTLPGIPLASQVPSWQVDSLPPQLQELPPIGLPQF----- 481
QY 411 QPPTAGVSVPPGFP-----PAPANISTACFSPGVPTAHSNTPTTQAPLLRE 459
Db 482 -----GLPPPGVGLSATVLRPOHSMPLNYGMPLFSAGFPFAYS-----SVPRSAISD 530
QY 460 EFPYRQNDKGRSKRPYSSGSSYSRSSYDSSQGLAQHIALTLSPSAHTLDDLHDHPH 519
Db 531 EWNAAFQNDKRS-----SRDHKDRTRRKDH----- 558
QY 520 PEEAEKRSAMTYHMDLMDIAHARSPPYRRTYRSRSPPEPFRGOSPTRKNVPREKER 579
Db 559 -----DSRSR-----RRDSSSSSSSMSSSSSDEDERRRRRKREKS 593
QY 580 EYFNRYREVPPYDIKAYYGRSVDFDPPEKERYREWEKRYEWEKYKGYAAGQAPR 639
Db 594 SKRRRSVEKERP-----RRADERRRDORDRDRDSDHD----- 629
QY 640 SANREDFSPERLLPLNIRNSPFTGRREDYAGQSHRNRLGANYPEKLSRDSSNAKDN 699
Db 630 -----VRS-----GRSKDIKASSHR-----RD-----RD 651
QY 700 PKSEKESENVPGDGNKKHKKRRRNEEGESESFLNPELETSRCKRSGSSGIDETK 759
Db 652 ARKDRRRRDV-----RKREKREKREED-----DDOK 679
QY 760 TPTLFLPLPSRDATPVPRDEPMDSITFKSVSDKDKREKDKPKVSKDKTRKRSKDSGATK 819

Db 680 T----- KOAESKDEIDID-----GIAE 699

QY 820 KNNVLKPSGPOEKVDGDEKSPREPLKAKAEKTKIDSVPSSSSQDEKVTGTPRK 879

Db 700 YGNV-----QADDTGDGAQ-----VADENAQND-----DSTSPKEEVSPPDE 739

QY 880 AHSKSAKDTRRSQPRTRRSKRTVPKTSQSQSPVPTRRPSLRKINYLARKNERE-- 937

Db 740 THEEDADE-----EPV-----DDEMTLEYV 759

QY 938 KKKKSYVDKDFESSMAKISKVEGTIYKPSPKRMEDVEKLETPP---KDKLAASST-- 992

Db 760 VKDKSLDPYQ--AMTSAEVAETKEESVPVEEDEPEHHDDVDVHKHKKKSQSTIN 817

QY 993 -----PAKKIKLNETGKKKIGNAENASTTKEPSEKLESTSSKIKOEKYGKGRKRVAG 1045

Db 818 TADEDEESSKMKKKHKKSKNNKHHKKEDGEDDEBRKKKKHKKHKKKKKKKKEND-- 874

QY 1046 SEGSSSTLVDTYSTSGSPVAKSEKTDYKRTVIKIMEYVNNNDTAPAEYIIMQVP 1105

Db 875 -----EDEDDDTTEKKKKKKKKYODDDDEDSNR-----KEP 906

QY 1106 OSKMWDKDFESEEDYKVTQPIQSVGKPSIIKN-VYTPPSAKYTKESKE-OPEKLOK 1163

Db 907 RLDDDEDKYELD-----KNFADKKEKKKKNNDEEDLTFEDRKEE 946

QY 1164 LPEKASHLMQHELRRSSKGSASSEKGRADREHS-GSEKDNPDKRSAGAPDKESTVDRL 1222

Db 947 LPEKS-----DRDRKRDRHDDDEDRKERHERHSQKIDEDDRKKERKRDRE----- 993

QY 1223 SBQGHFTLSQSKETRTSEKHSYVSGSSNKDPTPCRDKVDVDSRDYSSSKR--RDERG 1280

Db 994 -----FAYNSDAPKTYVKREDKRDROKDYERERKRRKDYEEKSKRSKSDRDNK 1046

QY 1281 ELARRKDSPPRKESISGOKSKLREERDLPKKG-----AESKKSNS-----PPDKK 1328

Db 1047 EKQREKEVEKEHEKDKRKEKRTIYEKSEKPRKSVHERMOKADSTSSSTYTPAPSLERK 1106

QY 1329 P-----HDHKAPYETRPCEETPAPVOKNSGKEREKHA 1360

Db 1107 PVSFTVASSKPTTNIRVROYSSSSSTKEDBEDERSKDRKKRDET---DYESIGEREKKS 1163

QY 1361 AEARNGKES 1369

Db 1164 SSRKVPKES 1172

RESULT 5

T34418

Cytothetical protein F12F3.3 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34418

R:Fulton, B.; Wohlmann, P.

Submitted to the EMBL Data Library, July 1998

A:Description: The sequence of *C. elegans* cosmid F12F3.

A:Reference number: Z21521

A:Accession: T34418

A:Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3488 <FOL>

A:Cross-references: EMBL:U80022; PIDN:AAC2585.1; GSPDB:GN00023; CESP:F12F3.3

C:Genetics:

A:Gene: CESP:F12F3.3

A:Map position: 5

A:introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

Query Match	5.2%;	Score 383;	DB 2;	Length 3488;
Best Local Similarity	21.3%;	Pred. No. 2.5e-07;		
Matches 213;	Conservative 183;	Mismatches 422;	Indels 184;	Gaps 41;

QY	464	EONKGRSEKFRPVSGSSYSSTSDSSGLA-----OHINATLTPSEAAHTLDLIDHP	517
Db	351	EKKDDGDDKSKPTTKKIIKKKETPSEEDVTAAPERBOKISVDYO-SVAT-----	401
QY	518	HPPEEAARSAMTYMHMDLMDIAHARSRPARYRYSRSRSPBPEFGOSPTRYVPREEK	577
Db	402	----EVCAGKKKPADEKP--PDLISKAKKD-----KSKKDPEASTEEKSTTEKPYNDK	449
QY	578	EREFENYRREUPRYDILKAYGSRVDFRDEPEKERYREMERKYRPMYKYYKGAVAOP	637
Db	450	TSKSAEKKTYKPKPEET---GKPLAKKPVEDKDAOSPPSSK-----SSP	495
QY	638	RPSANRDEFSBERLLPLNIRN---SEPT-----RCGRREDYAGOSHNRNLGN	663
Db	496	TDGAKKKQIPALFIPREIISRFGDPSJHSEJNTITTTIRG-REGSADAKPLVEPLSAS	554
QY	664	YPELSTRDSHNMKDNKSKSEKSESENVPPDDGKNKKHRRNRNKEGSEESTLNBELL	743
Db	555	VSMKVFTL-VESSAEKAEFSFKRSETPRDCKRKKKEGLRPAKSKSKKEVTAEKOSTEAL	613
QY	744	ETSRCKGSSGIDETKYDTLFTVPSRBDATPYRDEPMDAESITTFESYDKNKRE-KDKPK	802
Db	614	IESKK-----KEYDESKISE--QOPSDKNKSEVGVPEKAAGPEK---KDVSEIEVPK	663
QY	803	VKSOKTRKSDGSAITAKDNVILPKSGPOEKVDGDRKESPSSEPLPKAKAEAKIDSVK	862
Db	664	KRTIKKTEKSDSISISOKSVNLKPADDKSKSDVDYDS-----KTTEDDYKV---712	
QY	863	PSSSQDQDEVYGTGPRKASAKADTRRQSDPRTRRSKRTVPKTISSQKQFVRFRPRSL	922
Db	713	--ANDSKLEAADPTQIETEVVDDKSKKYLKAKKTEKS--DSFISQSETPRVEPETPK	769
QY	923	RKIVYLAREKNEREKRRK---SVDKDFSSSMKIS---KVEGEIYKPSPKMKMGDY	975
Db	770	AESPAOKIAEVNAKQOEVDONLKEAEVAARKLADLEKLTIEEMANKTKT-----AEV	823
QY	976	EKLERTPEPK-----KIASSTTPPAKKIKLIRETG-KKIGNAENASTYKPESEK-----	1022
Db	824	EAARKQEKEDQKLETEYVSVKSKSAEKELEBKQIKKAAEDAVVKKQKELNEKNKLEA	883
QY	1023	-LESTSSKIOEKYVKGKARKKAVGSGSSSTLYDVTSTSSJGSGSVKRSSEKT---DTRK	1078
Db	884	AKKAAOKLLEESAAKSKV-----SESVKAGEEKT-----KAGEKTYQVSESP	931
QY	1079	TVITMEEYNNMDYAPAEVDIIMIQVQPSKMPKODFESEEBDV--KTTOPLOSQSPSSI	1136
Db	932	TSKKTIDTQVGAATEPAD-----ETPKKIKIKKTEKSDSISQKSTADSEKVSQKE-	984
QY	1137	IKANTYTPKAT-----AKYTKESEBQEPKIOKLPKASHEMLQHOHLRSKSGASSEKG	1189
Db	985	-ODEPTTPAVSETOMYTEADRSKQKQETEKI-KIDAEIATAKTYQOADEDEKSKLDAOEKIK	1042
QY	1190	RAKREHRSGESEKDNPKRRKSGAO-PKRESTYV--RLSEQGHFKTLSSOSKETRTSEKHS	1246
Db	1043	KVSEDDAAREKELEINDKLTLESEIATRKASADKLTLEBOAQAANKAAEVAEAKQKEDQ	1102
QY	1247	VR-----GSSNK-----DFTTGRD--KKYDYDSROYSSSKRRDEGELARR	1285
Db	1103	LKIDTEAASKAAAEKLELEKQAIKKAAAGAVYKQKLEDEKNKLENNKNSAAGKLTIE	1162
QY	1286	KDSPRKEKESLSGO-----KSKLREER-DLPKGAESKSNSSPPRDKPK-----H	1330
Db	1163	EESAAKSKQVVEQAKLIDQYATAKTAKEKQTKLEKDEKSTKSESSEMYVDEKPKKKVLKK	1222
QY	1331	DHKAPYTKRCPCEYTPVDKNSG---KEKEKAADARNGKES	1369
Db	1223	TEKSDSSIQKSETSKTYVESAPSESEFYQYAADARQOKET	1264

RESULT 6
T13564
microtubule-associated protein homolog - fruit fly (*Drosophila melanogaster*)
M/Alternate names: hypothetical protein EG:49E4.1

C:Species: Drosophila melanogaster
 C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C:Accession: T13564
 R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
 A:Submitted to the EMBL Data Library, April 1999
 A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
 A:Reference number: 217689
 A:Accession: T13564
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-5327 <SPA>
 A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
 A:Genetics:
 A:Cross-references: FLYBase:FBgn0025392
 A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
 A:Note: EG:4984.1
 C:Superfamily: Drosophila 576k microtubule-associated protein homolog

Query Match 5.1%; Score 374.5; DB 2; Length 5327;
 Best Local Similarity 21.5%; Pred. No. 8.2e-07;
 Matches 317; Conservative 217; Mismatches 606; Indels 333; Gaps 69;

QY 19 KYAIPIDAEAYIAIGKKEKPPFLPEBSSSSSEED--DPIPAELLCLICKDITMDAV--- 72
 DB 983 KAEVODIATFAKKDIASR---TEEQALAKPAEELSSPTPEKLSKTSQDKKDOIGAPV 1038
 QY 73 -VIPCCGNSCDECIKRTTLESKDKHCTCPTQOND-----VSPDALLANKELROAVNFK 125
 DB 1039 DVLVNLQESIPPEKFKSTIESGATTAPTLPEDERIPLDQIKEDLVIEEKYKEE----- 1093
 QY 126 NETGYTRLRKQPLFLVPPRPPLSORNLQPKRSRPLNQOPVYRVRYVPTCSPTK 185
 DB 1094 -----TEEAIVAVVAVOTLPDPAALADTI-----LASATK 1125
 QY 186 TAGSCSDSGTSLRPL-----APSIISLTSSNOSSLAPVSGNPSPAPAPVDITATVSI 239
 DB 1126 DAPDANNAEALGELPDGSEVRLPMKMTPEAQONLLRDVTK-----TDEVA--DLP 1174
 QY 240 VHSKSDGPF-RDSDNKLLPAAALTSEHSKGASSIATATAMEKGV-----G 286
 DB 1175 VHEADADGLYEKDSQD-----ANAKSISHEKESK-----EKEETDEKENKVGIELG 1223
 QY 287 TSPNNSIFVGGSLHQQ-----LIPTGPIRINARCGGPGMEHKNKGLYLVSP 337
 DB 1224 DEP-NKVDISHVLKESVOEAEKVAVVETVEKKOEIVETATVITQENQEDLMEOVKD 1282
 QY 338 ----POQIRRG--ERSCYRSINGRHSHSERSQPTQSPSLPATPCFVVPPLVPPPH 390
 DB 1283 KEHEQKTESGIIITEKEKKKAST-PEEKETSDITSDELTA----- 1323
 QY 391 TLPLPPGVPPPOFSPOFES--SOPPT-----AGYSVPP-----GFP 426
 DB 1324 QLADPTTVPKASADREDTGSIESPRTIEALEVEVQAQKQAPKVPAPAEAIKTEKSP 1383
 QY 427 APANISTACFPGVPTAINTMPTTQAPLRSREFFYRQONKGRSKRPYSGSSISR-SS 485
 DB 1384 ASKETSRPESATGVKEDTEQTKSKKSPVSRPE-----SEAKDKKSPFASGEASRPES 1437
 QY 486 YTDSOGIAGHIHALTISPSAAH-----TLDLLHHPPEAEARSAMIVHMDLMDIAH 541
 DB 1438 VAEVKNQDGAESRRESIATHTKDESLD-----KAKQESR-----RESIAESIK 1484
 QY 542 ARSSSPYRRYRSRSPPEF--RGOSPTKRVNPREKEKEEYFNRYEVPPYDIKAVY 598
 DB 1485 PESGIDEKSALAKESASPEVTDKSKPESRESIAESILKAESTDKESAPPSKASRP 1544
 QY 599 GRVDFRPPFEKERYREMYEKYKYGAVGAOPRPSANEDSPERILPLNIRN 658
 DB 1545 SVESVVKD--LTEKSKESR-----ESIASAP-PIEFESVSRPSVID----- 1587
 QY 659 SPFTGRREDYAAQOSHNRNLG--GNYPKLTSTRDSHNAKDNP-KSEK-ESENVPGD 713

DB 1588 -----GIKDESAKPESSRRDSESLAKESAPESV-----LESVKDEPIKSTEKSRRESVAES 1638
 QY 714 GKGNKKHKKRRNRNDEKGESESEFLNPELLTSSKRCRSSGIDETKDTLVLPSRDAT 773
 DB 1639 FKAOSTK-----DEKSPLTGKDISRPE-----SAV-EVWMAFKETSRPSA 1680
 QY 774 --PYRDEPMAESITFKSVSD--KDKREKDKPKYSDKTKRKSQGSATAPKADNYLPSKG 829
 DB 1681 VGSKMDSEMSKEPSSRREYVKNGAQSRISRPASVAESA---DQADDLKE--LSRPEST 1735
 QY 830 POEYVGD--REKSPREPLPKAKEATKIDSVKPSSSQKDEKVTGTPRKHSKAKD 887
 DB 1736 TQSKREAGSIKDEKSP-----LASEASRPASV--AESVKDEA-----EKSKKE 1776
 QY 888 TRROSQPTTRSKRTVPTSSQKQPVTRRPRSLRKNYLIAREKNREKRSYVDKF 947
 DB 1777 SRRESVAESK-----PLPSKEASRPASV--AESIKD-----EAKSKESRRRESVAES 1823
 QY 948 ESSSMKISKVEGT--EIVKPPKRMKMGDVETL-ERTPEKDRIAS-STTPAKKIK---L 999
 DB 1824 PLPSKEASRPASVAESIKDEAKESKEESRRRESVAESKPLPSKEASRPASVAESIKDEAK 1883
 QY 1000 NRETGKRIAGNAENASTKPESEKLESTSSKIKQEKVKGAR-RKVAGSESSSTLVDTYS 1058
 DB 1884 SKESRRRESVAESKPLPSKEASRPASVAESIKDEAKESKEESRRRESVAESKPLPSKEASR 1943
 QY 1059 TSSGSGSPVRSSEKTDPTKRTVITMEYNNDNTPAADVLIIMIOVPS---KMDKDFE 1115
 DB 1944 PASVAESIKDEAKESKESR-----RESVAESKPLPSKEASRPASVAESIKDEAKESKEE 1998
 QY 1116 SEEDVY--KTQPIQSVKPSITIKVNTKPSAVPVKYEKESQEQEKLOKLP---KEASH 1170
 DB 1999 SRRESVAESKPLPSKEASRPASVAESIKDE-----AEKSKESRRRESVAESKPLPSKEASR 2054
 QY 1171 -----ELMOHELRSK-----GSASSEKGRADREHS-----GSEKDNDRKKSQAQDPK 1215
 DB 2055 PASVAESIKDEAKESKESRRRESVAESKPLPSKEASRPASVAESIKDEAKS-----E 2108
 QY 1216 ESTYDRLSEQGHFKTL---SOSKETRTSEKHESVRCSSKNDFTPGDKKVDYDSRYS 1271
 DB 2109 ESRRSMAESKQASIKQDQPLKVSRESVAESVKDDPKSKPSRRESV---AGSVT 2165
 QY 1272 SSKRDERGELARRKDPDPRGKESLSGQSKLREERDLPKGAESKKNSSPPRDKKPHD 1331
 DB 2166 ADSARDQSPLESKASRP---ESV---VDSYKDAEAOESRRRESKTSVILPPKAK---D 2216
 QY 1332 HKAPYETKRPCEETKPVVDKNSGKEREKHAABAR 1364
 DB 2217 DKSPREVLPVSMETIREDADQPKKPSQAESR 2249

RESULT 7
 T29757
 protein UNC-89 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
 C:Accession: T29757
 R:Du, Z.; Le, T.T.; Wilson, R.
 A:Submitted to the EMBL Data Library, May 1997
 A:Description: The sequence of C. elegans cosmid C09D1.
 A:Reference number: 220679
 A:Accession: T29757
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-6642 <DUZ>
 A:Cross-references: EMBL:AF003131; PIDN:AB54132.1; GSPDB:GN00019; CESP:unc-89
 A:Experimental source: strain Bristol N2; clone C09D1
 A:Genetics:
 A:Gene: CESP:unc-89
 A:Map position: 1
 A:Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1
 /3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Query Match	4.9%;	Score 363.5;	DB 2;	Length 6642;
Best Local Similarity	24.5%;	Pred. No. 2.8e-06;		
Matches 180;	Conservative 120;	Mismatches 260;	Indels 175;	Gaps 34

QY	701	KSKEKESN---VPDQKG-----	-----NKHKHKKRRRNEEKGESSESLN	739
Db	1304	ETKESESATTVIGGGSGVTEGSISVKT	IEVSKTDSQTDVREGTPKKRRVSAEELP	1363
QY	740	PELETSKCKGSSGIDTKTDTLFLVPSRD	---APFVRDEMDAESITFKSVSDOKR	796
Db	1364	KEVIDSDKKKSSPDKKES----	PEKTEKRPASPTKKTGEEVSPKPKSPASPTK	1418
QY	797	EKD--KPVKSDKTKRKSDGSATAKKDNV	LPSKPOQKRVGDRKESRSEP-----	846
Db	1419	EXSPAEEVKSPTKKEKSPSPPTKKEKSP	SPSTKKTGDEV---KEKSPKPTKKEKSP	1475
QY	847	-----PLAKAK-EATKIDSVKPBSSSQ	DEKXTGTPRKRAHKSADTROSOPTRR	898
Db	1476	KPEDYKSPVKKESKSPDANIVEVSETE	ITETMTT-----EMTESESESRISV	1526
QY	899	SKRPVTKSSOKSOPVTRPRPSLKI	NYLAREKNEBKRRKSVDRFESSMKISVE	958
Db	1527	KKEKTPKVEDEPKS-PTKKDKSPK---	SITEELKSVMKKEKSEKVEEKPASTTKK	1582
QY	959	GTEIVKP--SPKRMKGDE---KLERPE	KDIASSITTPAKKIKILNRETGRKIGNAENAS	1014
Db	1583	SPE--KPAASPTKKSENEVKSPTKKE	KSEKSVBELKSPKEK-----SPEKADDKP	1634
QY	1015	TTKEPS-EK--LESTSSIKOEKVGAK	KRVKAGSESSSLVDYTSSTSGSGSVRKE	1071
Db	1635	TKKEKSPKSAIEDVKSPTKKEKSEK	VEEKP-----TSPTKKESSPTKKT	1681
QY	1072	E--KTDKTRJY-KTMEYNDNDNTAP	ADVDYIMIQVQSKMDKODFEEDEEDVKTQ---	1125
Db	1682	DEVKSPTKKESPOVEE-----KPA	S-----PTKKESEKSVEEVYKPKPKS	1726
QY	1126	PIOSVGKSSIIKNVTPRSATAKYTE	KESOEPEKLOKLPKREASHELMOHELRRSSKGSAS	1185
Db	1727	PEKAEERP-----KSPTKKEKSP	EKSAEEVKSPTKKEKSEPK-----SA	1766
QY	1166	SEKGAOKREHSGSKDNPKRKSAG	ODPKDESTVURLSEQHFKLLOSSEKTRTSEKH-	1244
Db	1767	EKPSPKPKKESPPKAMADDEVKSPT	K-----KEKSPKEV---KPA	1819
QY	1245	ESVSGSSNKDPTPCROKVDVDSRD	YSSSKRRDRGELARKRDSPPGKESLSGOKSL	1303
Db	1820	AELLSPTPKKESPSPTKKTGDE	SEKSEPEKPEKPKSPPTKPKSP-----	1866
QY	1304	REERDLPKKGSEKSSNGSPRDK	KPHDHKAJETTKRCOEETKPYDKNSGKERE---	KHA 1360
Db	1867	-----GSPKKKKSPEAKRP---	APKLRLDKLOQ--VNKTDLAHFVEVVEHA	1911
QY	1361	AEAR---NGKESGA 1372		
Db	1912	TECKWFLDGCKEITTA 1926		

RESULT 8
151618
nucleolar phosphoprotein - African clawed frog
C:Species: *Xenopus laevis* (African clawed frog)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999
C:Accession: 151618; S57757
R:Calins, C.; McStay, B.
J. Cell Sci. 108, 3339-3347, 1995
A:Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp180
A:Reference number: 151618; MUID:96019267
A:Accession: 151618
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-990 <CAB>
A:Cross-references: EMBL:X88927; NTD:9895920; PIDD:CAAC1368.1; PTD:9895921

C:Genetics:
A:Gene: xNopp180
C:superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C:keywords: phosphoprotein

Query Match	4.7%;	Score 349;	DB 2;	Length 990;
Best Local Similarity	22.5%;	Pred. No. 1.4e-06;		
Matches 201;	Conservative 114;	Mismatches 420;	Indels 158;	Gaps 33;

[illegible]

RESULT
B47328

natural killer cell tumor-recognition protein - mouse
N:Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tum
C:Species: Mus musculus (house mouse)
C>Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 05-Nov-1999
C:Accession: B47328; I77662
R:Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993
A:Title: A cyclophilin-related protein involved in the function of natural killer cells.
A:Reference number: A47328; MUID:93135824
A:Accession: B47328
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-1507 <AND>
A:Cross-references: GB:I04289; NID:g192866
A>Note: authors translated the codon AGT for residue 972 as Arg
R:Rinfret, A.; Anderson, S.K.
Mol. Immunol. 30, 1307-1313, 1993
A:Title: IL-2 regulates the expression of the NK-TR gene via an alternate RNA splicing m
A:Reference number: I57820; MUID:94019422
A:Accession: I77662
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 235-237;263-294 <RIN>
A:Cross-references: GB:S65998; NID:g425701; PIDN:AAB28500.1; PID:g425702
C:Genetics:
A:Gene: NK-TR
C:Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology
C:Keywords: alternative splicing; lymphocyte
F:60-230/Domain: cyclophilin homology <CYP>

Query Match 4.6%; Score 342; DB 2; Length 1507;
Best Local Similarity 20.1%; Pred. No. 4e-06;
Matches 304; Conservative 186; Mismatches 569; Indels 456; Gaps 65;

65 KDIIMDAVIPCNGNSCDECIPTLLESD---KHTCPTCQ-----NDVSPD 109
Db 208 EULKIDASRPADRVIDCGVLAIRKIDVEPKRRKPTCEGSDSSRSSSSSE 267
Qy 110 ALIANKFLRQAVNNKNETGYTKRLKOLPPLFLVPPRLPSQRLNP----- 158
Db 268 SEVEETITRRRRHRRPKVRAKRRKREMS-----SSEPRKRTVSPEGYSERSDVE 321
Qy 159 -RSRPILRQODPVYFRTVSPCTDTAGSCSDSGTSLRPAISIS-SLTSNOSLAP 216
Db 322 KRSVDSNTRKRPVVPPEIPV-----PENRFLRDMPAITVPEPQNIPIVAP 371
Qy 217 PVSGNPSSAPAVPDITATVSIVHS---EKSDGPFRODNKLLP-----AA 260
Db 372 VVSDDKPSVSKSGRKIKRGITRIYHTPPRSRSHSESKDDSDSETPPHKEEMQRLRAYRP 431
Qy 261 ALTSHSKGASSIAITLMEKGVPTSPWNSIFVQ----- 297
Db 432 PSGEKWSKG-----DKLSDPCSSRWDERSLSQRSWSYNGYSDLSTARSHDGH 481
Qy 298 -----SLNLGOLPTTGPR-----INARPG 320
Db 482 HKHKRKKFKHKRKAOKHCRHRHQTKKRIIVMPDLSPSPHHRMKSQCVRRERSRA 541
Qy 321 GRPGEHNSKILGYLVSPPOQ---IRGERSCYRSINRGHNSERSORTO----- 366
Db 542 SSSSHSHSKRMWSKSDODDGSASTHSSRDSYRSKSHSDSGSRGSRGAAYAKSSRSIN 601
Qy 367 -----SPSLPATPCFVPPPPPLYP-PPHTLLPPEGV-----PPOFSPOFP-- 408
Db 602 RSKSRSSSRSGPRRTSISPKPAQLSENKPVTEPLRPVONGVILVQVAAENIPVIR 661
Qy 409 -SSOPTAGSYVPPGPPAPANISTACSPGVPTAHSTMTQTQALISREFFREQND 467
Db 662 LSDSPPSKWK---PGQRP-----WKPSYERIQEKAKATTN--LLPVOSTYSLNVI 707
Qy 468 KGRSKFPY-----SGSSYSRSSYDSSQGLAQHIALTLSPSAHITLIDLHDPH 518

Db 708 KATVSSSYHKRKEPSESDGSAYSK--YSDRSSG----- 739
Qy 519 PPEEASNSAMIVHMPDLMDIAHARSPPYTKRYSRSRSPPEFNGQSTKRNVPREKE 578
Db 740 -----SSGRSGS-----KSSRSRSSRSYTSRSRS-----LPTSLSLSRSS 778
Qy 579 REYF-NRREVPYPIDIKAY-----GRSVFROPFEK-----ERYEMERKYNEWYEK 626
Db 779 RSHSNKTSIDGQSHRSSSYTSVSSDDGRAMFRSNRKKSVTSKRRHNSNKK--TLASK 836
Qy 627 YKGYAVGAQPRPSANRDEPSPERLLPLNRSPTRGARE-DYA--AGQSH----- 675
Db 837 YVRG-----REKSS-----RRKXSSSSLDYTSDDSQSHVQYYSAP 875
Qy 676 -----RNRNLGNYPE-----KLSTRSHNAKNPNKSKKESEVPGDG 714
Db 876 KEKQKQVEALNDKQKGRGEGKPEWECPRSKKENSEDH--SRDVSFGKKGACAGSKMDS 934
Qy 715 KGN-----KHKHKKRRNEKE--ESPSFLNPELLETSRKCRGSSGIDETKDTPLFVLP 767
Db 935 ESNSEQDYTKSRKSDPRKRGSEKEGEASDSESVGQSHITAK-----PPAKPTSTPLP 989
Qy 768 SRDDATPVDEPMDAESITFKSVS-----DKDRKDKDPKYSKTSKSDGSATAK 820
Db 990 GSDGAWKSR-RQSSASSESSCSNLGNIRGEPQOKHKSODLKDGHTRKAREKS-KAKK 1047
Qy 821 DNVLKPSKQPOKVDGDRKSPRSEPPPLKAKAEATKIDSVKSSSSOKDEKVTGTPR-K 879
Db 1048 D---KHKRKP-----RKQAFHMQPLEFGDDEEMNG-----KVYTDQDPK 1088
Qy 880 ASKSAKQTRRSQOPRTKRSKTYPTKTSQKQAPTRPRRLKINILIAEKNERER 939
Db 1089 RIVSEKCEVAKGIP-----NVEKTCDEGSSPSKP-----KTLLED 1126
Qy 940 KRSVDESSSMKISKEGTEIVRSPKRMK-----GDVEKLETPPEKDIAS--STTP 993
Db 1127 PLAEGBHPSSCPAPLKVAVENASSPQAQNHLEHNGPGGGEVLQTDMMELCTPDRIISP 1186
Qy 994 AKKIKLNRETGKIGNAENASTTKPESEKLESTSKIKOEKYGAKRRVAGSESS--- 1050
Db 1187 AK-----GEVVSPLAN-----HRLDSPVNIIPEDQECMAHPRAGGEOESSMS 1230
Qy 1051 -----STLVDTSMSSGSGPVRSKSEKTDKRVYIKMEY-----NNDNTA 1093
Db 1231 SKTLBESGVKODSSTSVT--SPVETSGKKEGAKCOMLTDWKPLQGVNLSVSTATTS 1288
Qy 1094 PAEDYIIMIOVPOSKWMDKDFESEEDVYKTPQIOVGKPSIIKVVTKPSATAKYTEK 1153
Db 1289 SALDVKALSTVEYK-----PQGLIEIKSKNV-----RPSGLPDEVKRTAFLNRPRNQ 1339
Qy 1154 ESEQPEKLOKLPKEASHLMQHELRRSSKGSASSEKRAKDRHSGSEKDNPKRSGAOP 1213
Db 1340 ESSSDQGTFRSGDD-----QSRSHSRKRSKSETKSHHRTYSYSHSRSRSSSTSS 1392
Qy 1214 DKESIVDRLSEQGHF-----KTLSSQSK-----ETPRSEKHSYRGSNNDFPGRKTKYD 1265
Db 1393 YRSRSTYSRSDKWSYSRGTRSRSSSYGSHHRTYSRKS--RSSYDL-HSRSRSTY 1449
Qy 1266 DSRDYSSSKRRDERELARKKDSPPRGKESLGGQSKLLEEDLPPKGAESKSSSPPR 1325
Db 1450 DSY-YSRSRSR-----QGRSDSYHRG-----RSYNNRSMGSGSYGSDSS-- 1490
Qy 1326 DKPHDHKAPYETKR 1340
Db 1491 DRSYSHRSPSESSR 1505

RESULT 10
A44265
trithorax homolog HTX, version 2 - human
N:Alternate names: acute lymphoblastic leukemia gene 1 protein (ALL-1)
C:Species: Homo sapiens (man)
C>Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 01-Dec-2000

C:Accession: A44265; A44264; I58112; I37165; I38485
R: Tkachuk, D.C.; Kohler, S.; Cleary, M.L.
Cell 71, 691-700, 1992
A:Title: Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocation
A:Reference number: A44265; MIM:93046667
A:Accession: A44265
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-3968 <TKA>
A:Cross-references: GB:L04284; NID:g184393; PIDN:AA58669.1; PID:g184394
R:Note: sequence extracted from NCBI backbone (NCBI:P117729)
R:Gu, Y.; Nakamura, T.; Alder, H.; Prasad, R.; Cnaanani, O.; Cimino, G.; Croce, C.M.; Can
Cell 71, 701-708, 1992
A:Title: The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 g
A:Reference number: A44264; MIM:93046668
A:Accession: A44264
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 63-316; GLINSLEK', 327, 'Q', 329, 'VR', 332, 'DKEGTPP', 340, 'T', 342, 'EDKTIVROSPRH
546, 'LQIESTSP', 2555-3554, 'N', 3556-3594, 'V', 3596-3899, 'A', 3901-3968 <GU1>
A:Cross-references: GB:L04731; NID:g339921
A:Note: sequence extracted from NCBI backbone (NCBI:P117779)
R:Djablji, M.; Sellieri, L.; Parry, P.; Bower, M.; Young, B.D.; Evans, G.A.
Nature Genet. 2, 113-118, 1992
A:Title: A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute
A:Reference number: I58112; MIM:93265134
A:Accession: I58112
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1317-1700, 'DD', 1703-1936, 'H', 1938-2180, 'S', 2182-2328 <DUA>
A:Cross-references: EMBL:X83604; NID:g897757; PIDN:CAA58584.1; PID:g899268
R:Marshall, R.; Grell, J.; Lochner, K.; Nilsson, I.; Steigler, G.; Zweckbrunner, I.; Beg
Br. J. Haematol. 90, 308-320, 1995
A:Title: Molecular analysis of the chromosomal breakpoint and fusion transcripts in the
A:Reference number: I37165; MIM:95315013
A:Accession: I37165
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1212-1603, 'GTE' <MAR>
A:Cross-references: EMBL:X83604; NID:g897757; PIDN:CAA58584.1; PID:g899268
R:Note: submitted to the EMBL/GenBank/DBJ databases by R. Marshall, 20 December 1994
R:Gu, Y.; Alder, H.; Nakamura, T.; Schichman, S.A.; Prasad, R.; Cnaanani, O.; Salto, H.;
Cancer Res. 54, 2327-2330, 1994
A:Title: Sequence analysis of the breakpoint cluster region in the ALL-1 gene involved in
A:Reference number: I38485
A:Accession: I38485
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1251-1486, 'G', 1488-1538 <RES>
A:Cross-references: EMBL:U04737; NID:g451554; PID:g451555
C:Genetics:
A:Gene: GDB:ML; HTX; ALL-1; HRX
A:Cross-references: GDB:128819; OMIM:159555
A:Map position: 11q23-11q23
A:Introns: 1338/1; 1362/3; 1406/3; 1444/3; 1493/3; 1525/3; 1566/1
A:Note: the list of introns is incomplete
C:Keywords: acute lymphoblastic leukemia; proto-oncogene; zinc finger
F:1434-1456/Region: zinc finger CCHC motif
F:1479-1506/Region: zinc finger CCHC motif
F:1527-1556/Region: zinc finger CCHC motif
F:1569-1586/Region: zinc finger CCHC motif
F:1873-1900/Region: zinc finger CCHC motif
F:1933-1955/Region: zinc finger CCHC motif

Query Match 4.6%; Score 341; DB 2; Length 3968;
Best Local Similarity 20.1%; Pred. No. 1.2e-05;
Matches 283; Conservative 153; Mismatches 463; Indels 512; Gaps 60;

QY 111 LINKFLROAVNNFKNETGTVTKRLRKQLPPF-----LPLVPPRPRLSQNLQ 157
DB 371 LLSNSTRGO-----KKGAKKRIEKAALQGRKVKYTKOVKNITRIPVAVSAISSRIK 423

QY 158 PRSRSPILROQDPVY-----FRYVSPICSDTKTAGSCSD-----SGTLRLPAP 202
DB 424 TPRFIEDDEYDPIKARLESTPNSRFS--APSQSSSEKSAASQHSQSSSSSSSP 482
QY 203 SISLSINSSQSS-----LAPVSGNPSSAPAPVDITATVTSVHSEKSDPFPDSOKKL 257
DB 483 SVDTSDSQASSEITQVLPERSSTPEVHP-PLP-----ISQSPNESNDRRSRYS---- 532
QY 258 PAAALTEHSHKSGASSITAITAMEKGVGTSPWNISIFVGOSLHGOLIPYTGVRIMAR 317
DB 533 -----VSERFSGRT----- 542
QY 318 PGGRPEWESHKSLKGLVYSPPODIRGERSQYNSINGRHSHRSQRTQSPSLPATPCFV 377
DB 543 -----TKRLSTLQSAPOQ-----QYSSSP----- 561
QY 378 PVPPEPLXPPEP-----HTLPLPGVYP--PQSPQFPSSQPTTAYS--VPPEG 423
DB 562 --PPPLTTPPLPQPPASSISDHT--FWLMPPTIPLASPLPASTAPMOGKRKRSILEPT 616
QY 424 F-----PPAPANISTACFS-PGV--PYAHSNTMPTTQAPLLSREFFREQNDKGRES 472
DB 617 FRMTSLKHSRSEPOYESSAKYAKBGLRKPIFDFNRPPLTP-----EDVGFAS 665
QY 473 KPPYSGSSYSRSSYTDSSQGLAHIALVSPSAHTLDLHDHPPEAEARSAMIVH 532
DB 666 GFASGTAASARFSPPLHSGTRFDMH--KRSP-----LLRAPRFPPEAHSRIESTSV 716
QY 533 MPRLMDIAHARSPPYRARRSRSPPERGOSPTRKRVPREREKERYNRRREVPPY 592
DB 717 LPNNFTSAGTSSGVSNNRKKRVFSPINSEPPSPS--HSMKRISGHLSSSELSPPLPPS 774
QY 593 DIRAAYGRSYD-----FRDPEKERYREMERKYREMYKGYGVAGQAPR--- 639
DB 775 SVSSLSISVPLATSAIANTPTFFPS-----HSLTQSGEAEKNQRPKQ 819
QY 640 -SANREDF---SPERLLPILNRSPTTGRREDYAAQSHRNRLNGANTPEKIS--TROSH 694
DB 820 TSAPEAFSSSSPTPLFPWFTPSCOTERGNKDKA-----PEELSKRDAD 865
QY 695 NAKNDPKRSKESENVPGDGKGNKKHKKRRNDEKGESESLNP-----ELL 743
DB 866 KSEYKDSRDRDRER---EKERRSRKRRKRGSEISQSSSALYPRGYSKRVVGEVY 921
QY 744 ETGRKSGSGSIEGTITDITLVLPSPRDATPVDEPDMAESIFKYSQDKREKPKV 803
DB 922 ATSSSAKKATGRKSS-----SHDSGT-----DITSVL----- 950
QY 804 KSDTKRKSDSATAKKDNVL-KPSKGPOEKVDGDRKSPRSEPLKAKE---EATKID 859
DB 951 -----GDTTAVYTKILIKKRGKNTLEKTYND-LGP-TAPSLKEKTKLCLSTPSSS 997
QY 860 SVKPSSSS-----QKDEKVTGTRKRAHSAKOTRQSOQPRFRSRKRYPKTS 907
DB 998 TVVHSTSSIGSMLAQADKLPMTRKRAVSLTKAKKAQDCKTEKRSKQYDQ-----PKAQ 1052
QY 908 SOKSOPRT--RRPRSLRKINLYLAREKNRERKRSKVDKDFSSSMKISKVGTETIVP 965
DB 1053 GQSDSDSEISVGRPR---IKH-VCRRAVALOKRAVAPD----- 1088
QY 966 SPARKMEGVLEKLEPP--EKDKIASSTPPAKIKILNRETGKKIIGNAENASTTKEPSEKL 1023
DB 1089 -----DMPTLSALPWEEREKILSS-----MGNDKSSIASGSEDEPLAPIKPI 1132
QY 1024 ES-TSSKIKQEK--VGKAKRKAAGSEB-----SSSTLVDTYTSSTSGSPVAKKSEKD 1075
DB 1133 KPVTRKKAPEDEPVYKGRGRRCGCPGCVPEDCVCTNCDKPKFGGKNTIKQDCK-- 1190
QY 1076 TKRTVILKTEEYNNNDTAPEDVYIIMIOVPOSKMDKDDFESEEDVKTQPIQSVKPPSS 1135
DB 1191 -----MRKCONLQMPKATLOKAKAVKKKKKSKTSSEKD-----SKESS 1232
QY 1136 IIKNVTTKPSATAKYTEKESEQPEKLOLPRKASHLMQHEILRSSKGSASSEKGRADRE 1195

Db 1233 YKKNVY-----DSSQKP----- 1244
QY 1196 HSGSKNDPDKKSGADPDKESTVDRLSQGHFKTLSSQSKTKRTSEKHESVGRSSNKDF 1255
Db 1245 -TPSAREDDPAPKSSSGSEPPRPKPVEEKSEEGNVASAPGPESSKATTPPASHKSSKQVSOQAL 1303
QY 1256 -----TPGRKKKVDYDSDRSDSSSKRNDGELARRKNDSPRGKESLSQSKLREED 1308
Db 1304 VTPPPPTTGPPRK-----EVKTTTPSEKKKOP-----P 1333
QY 1309 LPKGAESKSSNSPPRDKKPHDKAPYETKRPCEETKPVDK-----NS 1352
Db 1334 PPSGPEQSKOKKAPRPSIPYKQK-PKEKEKP-----PPVKNQENAGTLNLTSLNNGNS 1388
QY 1353 GKER-----EKHAFAKNGKESG 1371
Db 1389 SKOKIPADGVHRIYDFKEDCEANVWEMG 1419

RESULT 11
E88320
Protein F07A11.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: E88320
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C. ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: E88320
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2738 <STO>
A:Cross-references: GB:chr_11; PIDN:CAA93781.1; PID:g3881547; GSPDB:GN00020; CESP:F07A11
C:Genetics:
A:Gene: F07A11.6
A:Map position: 2

Query Match 4.6%; Score 340; DB 2; Length 2738;
Best Local Similarity 19.6%; Pred. No. 8.9e-06;
Matches 312; Conservative 194; Mismatches 610; Indels 474; Gaps 68;

QY 26 DAEVAIGKKERPPLEPPSSSEEDPIPAELICLICKIMTDAVIT-----PCCG 78
Db 421 DOENNASGSDSSSDSEEGSSSSNEDSDQNDV-----DEDEDEDVYSEKRRHEPEEG 474
QY 79 NSSC-----DECIRTTLESKHTCPTCHONDVSPDALIANKFLQAVN 122
Db 475 KSSSGNGHRDSDNDKDHEDSERFSOPSTSSH-ETSHSEPKOSEAYOSNF-SPLN 531
QY 123 NEKNETGY-----TKRLRKLOLPF-----LFLVPPRLSOR----- 154
Db 532 YQSGPQVEFLSKELKEFSPSTSSASSSDELEMEMDNLTRLLEMHMRPFIYVS 591
QY 155 -----NLQPRRS-----PILRQO-----DPVFRYVSPSCST 184
Db 592 FVNRIDETVELNQKARASYEKFTGRPFKCNNDDEVLTOKIVFHEPRDYYYEENP-CSEL 650
QY 189 KTA-----GSCSDGTLRLPAPSSISLTNOSSLAPRVGNSAPAPVPIATVYSIV 240
Db 651 EYRIDMKRLSTADLDFRAIDSKELGRQO-----PAGGRTSGRPSLDESRTNRLSF-- 703
QY 241 HSEKSDGPRDSDNKLPAALTLTSEHSGASSIATLALMEKGVPTSPWNSIFYGOSLL 300
Db 704 -----DSTHHPRELAQORSH-----SLCIGPM-----TPSTP----- 729
QY 301 HQGLPTTPPVYINAR-FGGGRPGWE-----HSNKLGYLVSPPOQIRGERSCTYRSTN 353
Db 730 -----PPTSQPLLVNTHTLPGTSQSTSGITTPRSSQPPPLMSPVS--RHNSMS-----S 778

QY 354 RGRHHSERSQRTQSPSLATPCFVPPVPPPLXPPRPHLLPLPGVPPQFSPQ--FPSQ 411
Db 779 TGRPASIQTLRHQSVMEFPD---VSIPEPL-PPYHDEMARGRTPPSRSSETWPLRS 834
QY 412 PPTAGYVPPPGFPAPANISTACSPGVPAHSTMPTQAPLRSREEFYREQNDKGR 471
Db 835 PPF-----GPIQNLMLTPIVPPPHL-----AAT 859
QY 472 SKFPYSGSSYSSTSDSOGIAQIHALTLSPSAHNTLDLHDPHP---PEEAANSA 528
Db 860 STGTHSVSSASHSTPRHSISGTPVH-----CEPSNKT---SQPPTKSRPEKVOIHD 910
QY 529 MYVHM--PDLMDIARHSPPYRRYRSRSRPPER-----GQSTP--KRVNP 573
Db 911 TTSKSGPSMAIALQASQSMTSQDPKASAPSTPVVRDAGSDLVQAINQNLGLRLP 970
QY 574 REEKEREYFNRYREVPVPY-----DIKAYGRSVDFDPPEKERYEMEKRYEMYEK 626
Db 971 RIEKSSALQNIQNHQPPRSMANSTPSTSTHQAMFKRKEKKEKKEERE--- 1026
QY 627 YKGYAVGAQPPRANREDESPERLLPLNRSPTTGRREDYAAQSHRNRLGNPPE 686
Db 1027 -----FEARREMKRKETKEER---NKRKEMERAKLEDE---EROERKR-----E 1064
QY 687 KLTSDSHNAKPNPSKKESEENVPGDCKGNHKKHKKRKRNEKEGESRFLNPEL-LET 745
Db 1065 KKEKDE-----RKKEKRYKKAKEKELKKHKKR---GDSSESDSDNDELIDLIV 1114
QY 746 SKRCGSSGIDETKTDTLVLRSDATPVREDPMAESITFKSVSDKKRREKDPKYS 805
Db 1115 RSTJEMT--QEKHQHALLLSKGIIE-----NLKSRRSRQKRAHDSKEKQ 1161
QY 806 DKTKKR-----SDGSATAKDNVLPKSKGPOEVGCDREKSPSEPPLLKAKEEATK 857
Db 1162 QKSQORRYLVIESDDEGGDKGNSNGEEDSEKAD--LPPPAPELSSEADQRLK 1218
QY 858 IDSVPRSSSGKDEKVTGTPRAHSAKAK-DPRRQSPPTRRSKR-----TVPKTS9Q 910
Db 1219 V-----LKEREGELTSSDDEDDHNDAGETHQORLTEDPENKROKSLTAYSSDEGEK 1273
QY 911 SOPVTRRPRS-----LRKINYLAREKNEREKRSYVDKDF--- 947
Db 1274 NVPKRRRSDSDDAAKHFGWGAKDQOKRKRLEH---RSSSEDSKNAKRDROIP 1329
QY 948 -----ESSMKISKVDEGTIVAPSPRKAGDVEKLEKPE----- 983
Db 1330 HEDVSDDEETEDGSRSRQSTSTISNTAKE-----RKEKSGKTPLRIVPEPTGPL 1382
QY 984 -KQTIASSTTPAKKITKLNETGKIGMAENASTKPESEKLESTSS-----KIRO 1032
Db 1383 LSPKILSPKHLSPKJSTSTSKRSSISDHENLISPRORANTTSTSTATTSKHEALSTPE 1442
QY 1033 EKVKGAKRRKAVAGSESSSTLVDTSTSGSPVYKSEKTDYKTKVTKTMEYNNMT 1092
Db 1443 KPLSPPVTAKSSVSTIDPSIRDEFMSAGADPM---STGAPVVLTKAAMKAFNS 1496
QY 1093 AAADY-----ITAIQVPQSMKXD-----DFESEEDVYKT 1124
Db 1497 TPKKVSYSLLIDCYMLGMAKNSSSGOHSSGSSSSSDGSGTSSDSDSDDEVPKOT 1556
QY 1125 OPIQSV-----GRPSTIKNVTTPKPSATK-----YTERESE-QPKLQKLPKEASH 1170
Db 1557 EYVTSIPVVASDNGSPENNVETPISVSTPBPPEFTTISEGSSSEPPAVBEP- EAS- 1614
QY 1171 ELMQHELRSSKGSASSSEKGRANDREHSGSEKDNPDKRKSGAOPDKESTVDRLSEOGHFYT 1230
Db 1615 --VEPQMETSONVPEVSEHEHDSHEHGDSE-----VAVSSQOQPLEHOREKEE 1660
QY 1231 LSQSKETRTSEKHESYRSSSKDFTPGDKKVDYDSDRSDSSKRDDEGELARRKDDSP 1290
Db 1661 LENKILDVAAEHHEVOG-----DEDSVSSSIAPSDDEDDPVTAQOEKSA 1706


```
Db 421 DQNNAGSDSSDSEEGSSSSNEDSDQNDY-----DEEDBDVYSEKKRHEPEEG 474
QY 79 NSGC-----DECIRTTLESKDHCTPTCHQNDVSDALIANFLQAAYN 122
Db 475 KSSSPGNHGRDESGDKDHEDSSERFSQPSSTSH-ETSHSPKDEAVQSRSF--SPLN 531
QY 123 NFKENGTGY--TKRLKOLPPE-----LFLVPPRPPLSOR-----154
Db 532 YQSQSGYELESEKEIKQEPSPTTSSASSDLEDMEMPNPLTRMLERHMRPFIDVSS 531
QY 155 -----NLQPRRS-----PIRQO-----DPVRYTVSPQSDT 184
Db 592 FVNRIDEIVELNOKARASYEFTGRPEPKCNNDENVLSIQKIVHEBRDYXYENP--CSEL 650
QY 185 KTA-----GSGSDGTLRLPAPISITLSSOSSLAPVSGNPSSAPARVDITATVSY 240
Db 651 EVAIRDMRKLSDTADLDDEFRAIDSKELGRQO---PAGRTSGRPSLDESRTNRLSF-- 703
QY 241 HSEKSDGPFPSDNKLLPAAALTSEHSGKASITATLMEKGVPGTSPWNSIFVGQSL 300
Db 704 -----DSTHPRALAQSH-----SICIGPM-----TPSTP-----729
QY 301 HGQLPTTGPRVINAAR-PEGGRPME-----HSNKLGYLSPPOQIRRGERSYRSIN 353
Db 730 -----PFTSOPILVNTHTLPTSQSTSGITTPRSSQPPIMSPVS--RHNSMS-----S 778
QY 354 RGNHSEBGRORQSPSLPAPRCFVPPRPPLYPPTHLLPAGVPPROQSPQ--FPSSQ 411
Db 779 TGRPAISIQTRKHOSVMPRPD--VSIPRPPI-PPTHIDEMAPRGTPRSKSSSTWPLNS 834
QY 412 PPRAGYSVPPGPPRANISTACFSPVTAHNTMPTTQAOLSKNEEYRQNDKGR 471
Db 835 PEP-----GPIQLMLTPIVPPHLL-----AAT 859
QY 472 SKFPYSGSYSRSSYDSSGGLAOHIALTLSSAHTLDDLHDHP--PEAEARSA 528
Db 860 STGTHSVSSAHSTPRHSISGTPVH-----CEPSNSKT--SQPPPKSRPEKVOIRHD 910
QY 529 MIYHM--PDLMDIAHARSRPYRRTYRSRSRSPPEFR-----GQSP--KRNVP 573
Db 911 TIKSGPSNAINLQARSQMTSGDRKKSAPSTPYVRDAGSDLVQAIMSQMLGLRKL 970
QY 574 REEKERYENRYREVPY-----DIKAYGRSVDFRDPEFEKERYREMYEK 626
Db 971 RIKKSSALONIONHOPRHSNANSTPSTSTHOAMFKKEREKKKEKEERE-----1026
QY 627 YKGVAVGAQPRSANREDSPEPLRLPLNRNRPTRGRREDVAAQGSNHRNLGNYPE 686
Db 1027 -----REARRRDKRKETKEER--NRKEMERAKRLED--EKOERKR-----E 1064
QY 687 KLSTRDSHNAKDKPKSEKSENVPGDGKNKHKKRRRNEKGESESFLNEL-LET 745
Db 1065 KKRERDE-----RKKEKKYVKKAKKEKLLKKKHKK--GSSDSDSDSDNDELDDV 1114
QY 746 SRKCGSSGIDETKTTFLVLPDRDATPYRDEPMAESTTEFYSVDKDKREKDKPVKS 805
Db 1115 RKSTKEWT--QEEKDQLALLLKGII-----NLKSRRSRDRADHSEKMG 1161
QY 806 DKRRKR-----SDGSATAKKDNVLPKSGPQEKYDGRKESPRSEPLKAKKEATK 857
Db 1162 QKSOQRVLTLESSDDDEGKDGKNGSSNGEESSEKAD--LPPPAPELSAADRKL 1218
QY 858 IDSVKSSSSQDKVYGTPRKAKHSKAK-DTRROSQPTRRSKR-----TYPKTSSOK 910
Db 1219 V-----LKEKEKELITSSDDEHDNDAGELHQQRITEDRNRKRKOKSLTAYSSDEGERK 1273
QY 911 SQVTRRRRPS-----LRKINTYIAEKNEREKKKSVDKP-----947
Db 1274 NVPRKMRDSDSEDAAKHPGWSAKDKQOKRRKLEH-----RRSSESESKNAKNAKREDIP 1329
QY 948 -----ESSMKISKVEGTETLVKPSPKKMGVDYKLERTPE-----983
```

```
Db 1330 HEDVSDDEETEDGSRSRROSTSTISNTAKE-----RKESGKTPLRIVEPTGPEL 1382
QY 984 -KDKIASSTTPAKKIKLNBRETGKIGNAEMNASTKPESEKLESTSS-----KIKO 1032
Db 1383 LSPKILSPHLSKPTSTSTKSSISDHNELISPRORNTTSTSTATSSKHEALSTPE 1442
QY 1033 EYKVGAKRRKRVAGSESSSTLVDTYSTSSGSPVRKSEKTDTKRYIKT--MEEYNN 1089
Db 1443 KPLSPVYTAKSSVSIDDPSIRDEFMNGSADSPM-----STTGRPVLTKAAMKAFN- 1495
QY 1090 DNTAPADYITIMQVOSKMDKDF-----ESEEDYKTTQPIQSGKSSIIKNY 1141
Db 1496 -STPRKETTQAVQSIFFDEEADDEFQYDPFGISTNEKEVSGKDPNIT--KTEPLNNGH 1552
QY 1142 T-----KPSATAKYTES-----EOPKLOKLPKASHE--LMQHELRSKGSASSEK 1188
Db 1553 TDLTFSPSSAHASEKQSTKSEDMEDSELYMEKEVMEQVIAQEVAVHPSPMEBE 1612
QY 1189 GRAKREHSGSEKDNPKRRSGAQDPKESTVDRLSQGHFKTILSQSKET-----RTSE 1242
Db 1613 VKL--ETSPVPRKEEPIKMBE--SPROTPTPLDIS-----NNESQTPGAVNNHLHE 1659
QY 1243 KHESVR-----GSSNKDFTGGRDKKVDYDSRDYS-----SKRRDRGEL 1282
Db 1660 NHDAYQTPQLQOPASQVQAQSPRPVAPDSDQNGPVLVSOOSQSPSPMSQSDMAQNL 1719
QY 1283 -----ARRKDSPPRGKESLSGQSKLNEERDLPRKG-----1313
Db 1720 ILSKINDLAKLHNPEALAQATRGDCSIFQHLLLAQGNQMTPEMIQAKAFPA 1779
QY 1314 -----AESKSSNSPPRODKKPHDHKAPV-ETKRCEBEKTPVDKNGSKEREKFA 1360
Db 1780 OQENENANOMQAKMQOQTINKDRIKEOERYKMYEENRKEVEDREROKREERORLA 1839
QY 1361 A 1361
Db 1840 A 1840

RESULT 15
T30826
nascnt polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
M:alternate names: alpha-NAC protein
C:species: Mus musculus (house mouse)
C:date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C:accession: T30826
R:Yotov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A:title: Differential splicing -ln of a proline-rich exon converts alphaNAC into a mus
A:reference number: Z20889; M0ID:96312450
A:accession: T30826
A:status: preliminary; translated from GB/EMBL/DBJ
A:molecule type: DNA
A:residues: 1-2187 <YOT>
A:cross-references: EMBL:U48363; NID:q1666688; PID:q1666689; PIDN:AA018732.1
C:genetics:
A:gene: Naca
A:map position: 10
A:introns: 24/1, 1996/1, 2024/3, 2050/3, 2099/3, 2142/3, 2183/3
A:note: differential splicing converts alphaNAC into a tissue-specific DNA-binding ac
C:keywords: alternative splicing; DNA binding; transcription factor

Query Match 4.4%; Score 327; DB 2; Length 2187;
Best local Similarity 20.9%; Pred. No. 2, 2e-05;
Matches 334; Conservative 182; Mismatches 640; Indels 442; Gaps 77;

QY 8 NMKGAMLTNTGKAITPDAEAIAIKKE-----KPPLEPPSSSEEDPTPALLC 61
Db 480 NLPISALVNVG--ADVSPQAQGLPPTKDTTLOPLALAKESPSSQS-----ASSLE 529
QY 62 LICKDLMTD-----AVIPLCGGNSCDECIRTTLLESQKHCTPCCHQNDVSPDALIAN 114
```


Db 530 VLSEDTYKTKTGGPAPVPAIAGVA-----TTTSLRADSPPAVIRADSCVSPNTV--S 582
Qy 115 KELROAVANNFNKNETGYTKRLRKQLPPLFLVPPRPLSORNLOPQRSPIILQOOPVVER 174
Db 583 OPLKRSYVD-----PAMAPRTAKNTAPSTISPLV----- 611
Qy 175 YTVSPFCSDTKTAGSCDSGTLRLPAPSISLITSNOSGLAPPVSGNPSAPAPYDITA 234
Db 612 -----PLASE-----GCPVASSMAL--SPONASVSEFALALSPEI--PKSVFPDPPL-A 656
Qy 235 TVSIS-----VHSEKSDGPRDSDNKLPAALITSEHSKG-----ASSIATITAMEEK 282
Db 657 EISFENARKVDVAVSHMESSGSSROGH---PDASYT---AKGTVCCLADSSLDTSVASAK 709
Qy 283 GVPGSPMNSIFVGGSLHGLIPTTGPRINAARPG-----GRPGWHSNKLGLVSP 338
Db 710 G-----SALSGASSPLX-PLEVSFLPEAGLAVGPGK--SLNKL--PTPP 750
Qy 339 OQIRGERSCYRSINGRHHSERSQRTOSPPLPATPCFVPVP-----PPLYPP 388
Db 751 -----SSKGAPVPSTGAPSPKGAPIVPTSSISSKQVPAELILPS 790
Qy 389 PHILP-----LPQVPPPOPSQF-----PSSQPTAGTVP----- 420
Db 791 POKTPEEVTASRLISAVQSPKVDPIMSDVTPTSPKKTSAVAVPKDTSATLSLKVAVTSL 850
Qy 421 -PPGPPAPANISTACFSGVPTAHSN-----TWPTQAPLISKEEYREBN 466
Db 851 SPPKAPVAPSNAT-IVPTEIPTSLKNALAAATPKETLATSLPKVTSP--SPQKTPKSVS 907
Qy 467 DKGRSEKPPYSGSYSRSSYTDSSO-----GLAQHI-----HALTLPSA 506
Db 908 LKGAPAMTSKKTATELAASKDVSPQFPEKVEPLQHVPTSPKSPYSDTLGSLALSPPK 967
Qy 507 AHTLDLHDHPHPP---PEAEARSAIYAMPDLMDIAHAKRSPPYRKYRSRSPPEFR 563
Db 968 GPPATIAETPTYPKSPKSPKAPKTPATPSPE--GVTAVPLEIPPCSKKAPKTAAPKESS 1025
Qy 564 GOSPTRKRVNPREKEREVEFNRYREVP-----PYDIKAYGRSVDFRDPFEKERYREME 617
Db 1026 ATSSSKR-APKTAVS-----KEIIPSKGVTAVPLEISLPLEKETSATPGEKSASSP-K 1076
Qy 618 RKYREMYEKYYKGYAVGAQPRPSANREDFSPERLLPLINRNSPFRGRREDYAAGOSHNR 677
Db 1077 RSPK-----TAGPRETPPGGVTAVPPETSLP--PKETPQNAKTPWESLAASQKRS 1124
Qy 678 RNL-----GG-----NYPEK-----LSTRDSHNAKDNPK 701
Db 1125 PKTSVPKETPPGCVTAMPLEIPSAPQKAPKTAVPKOIPTPEDAVTILAGSPLSPKKASKT 1184
Qy 702 SKRESEENVPGG---KGNKHKKRKRNEKEGESEFLLNPELLETISRKCRGSSGIDE 757
Db 1185 AAPKEAPATPSGVIAVSGEISPSPKKTSKTAAPKENSATLPP---KRSPKTAAPKETPA 1241
Qy 758 TKTDFLFLVPSHDDATPVDEPDMAE---SITFKSVSDKDKREKDKPKVKSDDTKRKS 812
Db 1242 TSSBEGTAVPSISPP---PTPAKSGVYVLTJPKGAPNALAESPAK-KVPKTAPE 1296
Qy 813 DGSATAKDNVLKPSKGOEKVGDREKSPRSEPLKKAKEEATKIDSVKSSSSOKDEK 872
Db 1297 ETSTTP-----SPQKIP--KVAGPKFAS--ATPPSKKPTKTAV-----PKETSAPSEG 1340
Qy 873 VVG-----TPRKAHSKSAKDTRRSQPRTRRSKRT--VPKTSQSKSQPVTRRRPSLAK 924
Db 1341 VTAVPLEITPPSPRKA-----PKTAPKETPAPEGATTAPQI--PPSPRK 1385
Qy 925 IYVLIAREKNEBERKRRKSVDK---DFESSMKISKVEGTE--IVKPSPKR----- 969
Db 1386 GSKKAGSKETPTTPEBEGTATAPLEIPISSKKTSKMASPKETLVTPSSKKLSQTVGPKET 1445
Qy 970 KMEGVEKILERTPEKDKIASSTTPAKKIKLN--RETGKKIGNAENASTTKE-----P 1019
Db 1446 SLEGATAVPLEITPSHKKAPKTVDPKOVPLTPSPDAPPTL--AESPPSSPKKAPKTAAP 1503

Qy 1020 SEKLEST---SSKIKQEKYKAKARKVAGSEGSSTLVDTYSTSTGSGSPVRK---SEEXT 1074
Db 1504 SERVTTVPEPEKATPQKASGTTASKVPVPAETQEVAVSSRETPVPVPAVPVKNPSHKKT 1563
Qy 1075 DTKRTVITMEEYNNNDNTAPAEVDIIMIOVPOSKWDKDFESEEDVKTTOPIQGVKPS 1134
Db 1564 S-----KTIEL-----KEAPA-----TLPPSPTRKSPKIPSSKKAPRTSAREPPASPS 1606
Qy 1135 STIKNVTTKPSATKYTEKESQPEKLOK-----LPKE--ASHELMOHLRS-----SKGS 1183
Db 1607 --IKPVTTSLAQT-----PPSLQAPSTTIPEKNLAPAVLVPSSKPPAAPARAS 1655
Qy 1184 ASSEKGRADREHSGSEKONPDKRKGA-----OPDKESTVDRLSEQGHFTLSQS 1234
Db 1656 ASLSRATAPAPAPAKENATTIPSCKKAATETPIENTASLEGAPKETSETSVKIMSS 1715
Qy 1235 SKETRTSEKHESVRGS-----SNKDFTPGRDKKYVDYDSRDYSSSKRRDERGEL 1282
Db 1716 PPKKASSSKRASTLTPATTLPSLKEASVLSPTATSSGKDSHISPVSDACSTGTTTPOASEK 1775
Qy 1283 ARKDSPPRGKESLS--GQSKLREEROLPKKGASKSNSP----PRDKKPHDKKAPY 1336
Db 1776 LPSKGPATFTEMLAAPAPESALATTAPLOKSPGANSNSASSPKCOPDPSKR--DTKGLP 1833
Qy 1337 ETKRPEETKPYDKNSGKREKHAAEARNGKESGANC 1374
Db 1834 SAVALAPQIVPEKDTSK-----AIEFTLLVSPAKGSDC 1866

Search completed: September 12, 2002, 19:13:04
Job time: 6549 sec

Fri Sep 13 10:04:42 2002

us-09-811-045a-1.rpt

Page 14

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2002, 17:49:20 : Search time 25.52 seconds
(without alignments)
2130.182 Million cell updates/sec

Title: US-09-811-045a-1
Perfect score: 7374
Sequence: 1 MMEVXPNNKGMALTRTGKY.....WRRSWLLGRRRAPSSRNPS 1404

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwisProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	353	4.8	3969	1 HRX_HUMAN	Q03164 homo sapien
2	342	4.6	1453	1 NRCR_MOUSE	P30415 mus musculu
3	337.5	4.6	1020	1 NEH_HUMAN	P12036 homo sapien
4	333.5	4.5	2468	1 MAPB_HUMAN	P46821 homo sapien
5	326	4.4	2492	1 ATRX_HUMAN	P46100 homo sapien
6	321	4.4	1391	1 MST2_DROXY	Q08696 drosophila
7	321	4.4	2464	1 MAPB_MOUSE	P14873 mus musculu
8	317	4.3	1462	1 NRCR_HUMAN	P30414 homo sapien
9	315	4.3	633	1 MLH_TETTH	P40631 tetrahymena
10	315	4.3	3866	1 HRX_MOUSE	P55200 mus musculu
11	305	4.1	1822	1 ZAP3_HUMAN	P47550 homo sapien
12	302	4.1	2842	1 APC_RAT	P70478 rattus norv
13	299.5	4.1	2459	1 MAPB_RAT	P15205 rattus norv
14	299	4.1	2476	1 ATRX_MOUSE	O61687 mus musculu
15	294	4.0	699	1 NP14_HUMAN	O14978 homo sapien
16	291	3.9	1359	1 ATRX_CAEEL	O91760 caenorhabdi
17	290	3.9	2805	1 MAPA_HUMAN	P78559 homo sapien
18	288.5	3.9	2845	1 APC_MOUSE	O61315 mus musculu
19	286.5	3.9	1790	1 USOL_YEAST	P23186 saccharomyc
20	285	3.9	705	1 TRDN_RABIT	Q28820 oryctolagus
21	283.5	3.8	488	1 CYL2_BOVIN	Q28092 bos taurus
22	282.5	3.8	831	1 NPH_RAT	P16884 rattus norv
23	279	3.8	728	1 TRDN_HUMAN	Q10061 homo sapien
24	278.5	3.8	1189	1 YOH6_YEAST	P47035 saccharomyc
25	277.5	3.8	667	1 CYL1_BOVIN	P35662 bos taurus
26	276.5	3.7	1210	1 AP4_HUMAN	P51825 homo sapien
27	276.5	3.7	2142	1 BAT2_HUMAN	P46634 homo sapien
28	276	3.7	771	1 CALD_CHICK	P19246 mus musculu
29	275.5	3.7	872	1 S3B2_HUMAN	Q13435 homo sapien
30	274	3.7	2843	1 APC_HUMAN	P25054 homo sapien
31	273.5	3.7	704	1 NP14_RAT	P41777 rattus norv
32	273.5	3.7	1781	1 ARAQ_HUMAN	Q02952 homo sapien
33	273	3.7	2774	1 MAPA_RAT	P34926 rattus norv

34	272	3.7	1220	1 IFP2_HUMAN	O60841 homo sapien
35	271.5	3.7	1386	1 ZAP3_MOUSE	O91017 mus musculu
36	271.5	3.7	2116	1 MYS2_DICDI	O08799 dictyosteli
37	270.5	3.7	793	1 CALD_HUMAN	O05682 homo sapien
38	270.5	3.7	1233	1 YF16_YEAST	P43597 saccharomyc
39	269.5	3.7	598	1 CYL1_HUMAN	P35663 homo sapien
40	269	3.6	1253	1 DSPP_HUMAN	O9enz4 homo sapien
41	266	3.6	1087	1 NPH_MOUSE	P19246 mus musculu
42	264.5	3.6	1374	1 RNC_HUMAN	O9arr4 homo sapien
43	263.5	3.6	1130	1 YL17_CAEEL	O11102 caenorhabdi
44	262	3.6	1898	1 TRHY_HUMAN	O07283 homo sapien
45	261.5	3.5	1581	1 PPRB_HUMAN	O15648 h peroxisom

ALIGNMENTS

RESULT	ID	HRX_HUMAN	STANDARD:	PRT: 3969 AA
AC	Q03164	Q14845; Q16364; Q13743; Q13744; Q90MA3;		
DT	01-OCT-1993	(Rel. 27, Created)		
DT	01-NOV-1995	(Rel. 32, Last sequence update)		
DT	01-MAR-2002	(Rel. 41, Last annotation update)		
DE		Zinc finger protein HRX (ALL-1) (Tritorax-like protein).		
GN		MLL OR HRX OR ALL1 OR TRX1 OR HTRX.		
OS		Homo sapiens (Human).		
OC		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC		Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX		NCBI_TaxID=9606;		
RN	[1]			
RP		SEQUENCE FROM N.A.		
RX		MEDLINE=9304667; PubMed=1423624;		
RA		Tkachuk D.C., Kohler S., Cleary M.L.;		
RT		"Involvement of a homolog of Drosophila trithorax by 11q23		
RT		chromosomal translocations in acute leukaemias.";		
RL		Cell 71:691-700(1992).		
RN	[2]			
RP		SEQUENCE FROM N.A.		
RX		MEDLINE=9629053; PubMed=8703835;		
RA		Nilson I., Loechner K., Slegler G., Grell J., Beck J.D., Fey G.H.,		
RA		Marschalek R.;		
RT		"Exon/intron structure of the human ALL-1 (MLL) gene involved in		
RT		translocations to chromosomal region 11q23 and acute leukaemias.";		
RL		Br. J. Haematol. 93:966-972(1996).		
RN	[3]			
RP		SEQUENCE OF 1-1909 FROM N.A.		
RX		MEDLINE=93390935; PubMed=8378076;		
RA		Yanamoto K., Seto M., Komatsu H., Iida S., Akao Y., Kojima S.,		
RA		Kodera Y., Nakazawa S., Ariyoshi Y., Takahashi T., Ueda R.;		
RT		"Two distinct portions of LFG19/ENL at 19p13 are involved in t(11;19)		
RT		leukemia.";		
RL		Oncogene 8:2617-2625(1993).		
RN	[4]			
RP		SEQUENCE OF 1317-2328 FROM N.A.		
RC		TISSUE=Brain:		
RX		MEDLINE=93265134; PubMed=1303259;		
RA		Djabali M., Seiler L., Parry P., Bower M., Young B.D., Evans G.A.;		
RT		"A trithorax-like gene is interrupted by chromosome 11q23		
RT		translocations in acute leukaemias.";		
RL		Nat. Genet. 2:113-118(1992).		
RN	[5]			
RP		SEQUENCE OF 1251-1538 FROM N.A.		
RX		MEDLINE=94215165; PubMed=8162575;		
RA		Gu Y., Alder H., Nakamura T., Schlichtman S.A., Prasad R., Canaan O.,		
RA		Saito H., Croce C.M., Canaan E.;		
RT		"Sequence analysis of the breakpoint cluster region in the ALL-1 gene		
RT		involved in acute leukemia.";		
RL		Cancer Res. 54:2326-2330(1994).		
RN	[6]			
RP		SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).		
RX		MEDLINE=95322025; PubMed=7598802;		
RA		Mbangkollo D., Burnett R., McCabe N., Thirman M., Gill H., Yu H.,		

RA Rowley J.D., Diaz M.O.;
 RT "The human MLL gene: nucleotide sequence, homology to the Drosophila
 RT trx zinc-finger domain, and alternative splicing";
 RL DNA Cell Biol. 14:475-483(1995).
 RN [17]
 RP SEQUENCE OF 1212-1603 FROM N.A.
 RX MEDLINE=95315013; PubMed=7794749;
 RA Marschalek R., Grell J., Lochner K., Nilsson I., Slegler G.,
 RA Zweckbronner I., Beck J.D., Fey G.H.;
 RT "Molecular analysis of the chromosomal breakpoint and fusion
 RT transcripts in the acute lymphoblastic SEM cell line with chromosomal
 RT translocation t(4;11)",
 RL Br. J. Haematol. 90:308-320(1995).
 RN [8]
 RP SEQUENCE OF 1421-1540 FROM N.A.
 RX MEDLINE=94020842; PubMed=8414518;
 RA Forster A., Rabbitts T.H.;
 RT "A method for identifying genes within yeast artificial chromosomes:
 RT application to isolation of MLL fusion cDNAs from acute leukaemia
 RT translocations.";
 RL Oncogene 8:3157-3160(1993).
 RN [9]
 RP CHROMOSOMAL TRANSLOCATION WITH GAST.
 RX MEDLINE=20183971; PubMed=10706619;
 RA Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,
 RA Wilson R.B., Jones D.H., Addya K., Leonard D.G.B., Kushner B.H.,
 RA Williams T.M., Lange B.J., Felix C.A.;
 RT "Detection of leukemia-associated MLL-GAST translocation early during
 RT chemotherapy with DNA topoisomerase II inhibitors.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:2814-2819(2000).
 RN [10]
 RP CHROMOSOMAL TRANSLOCATION WITH AF3P21.
 RX MEDLINE=20115194; PubMed=10648423;
 RA Sano K., Hayakawa A., Piao J.-H., Kosaka Y., Nakamura H.;
 RT "Novel SH3 protein encoded by the AF3P21 gene is fused to the mixed
 RT lineage leukemia protein in a therapy-related leukemia with
 RT t(3;11)(p21;q23).";
 RL Blood 95:1066-1068(2000).
 CC -1- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: HEART, LUNG, BRAIN AND T AND B LYMPHOCYTES.
 CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
 CC T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND MLT1/ENL;
 CC T(4;11)(Q21;Q23) THAT INVOLVES MLL AND MLT2/AF4; T(9;11)(P22;Q23)
 CC THAT INVOLVES MLL AND MLT3/AF9; T(6;11)(Q27;Q23) THAT INVOLVES
 CC MLL AND MLT4/AF6; T(11;17)(Q23;Q21) THAT INVOLVES MLL AND
 CC MLT6/AF17; T(X;11)(Q13;Q23) THAT INVOLVES MLL AND MLT7/AFX1;
 CC T(10;11)(P12;Q23) THAT INVOLVES MLL AND MLT10/AF10;
 CC T(1;11)(Q21;Q23) THAT INVOLVES MLL AND AF1Q; T(11;19)(Q23;P13.3)
 CC THAT INVOLVES MLL AND ELN; T(11;19)(Q23;P23) THAT INVOLVES MLL
 CC AND GAST; and t(3;11)(p21;q23) that involves MLL and AF3P21.
 CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
 CC -1- SIMILARITY: CONTAINS 1 CXXC-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L04284; AAA58669.1; -;
 DR EMBL; Z69744; CAA93625.1; JOINED.
 DR EMBL; Z69745; CAA93625.1; JOINED.
 DR EMBL; Z69746; CAA93625.1; JOINED.
 DR EMBL; Z69747; CAA93625.1; JOINED.
 DR EMBL; Z69748; CAA93625.1; JOINED.
 DR EMBL; Z69749; CAA93625.1; JOINED.
 DR EMBL; Z69750; CAA93625.1; JOINED.

DR EMBL; Z69751; CAA93625.1; JOINED.
 DR EMBL; Z69752; CAA93625.1; JOINED.
 DR EMBL; Z69753; CAA93625.1; JOINED.
 DR EMBL; Z69754; CAA93625.1; JOINED.
 DR EMBL; Z69755; CAA93625.1; JOINED.
 DR EMBL; Z69756; CAA93625.1; JOINED.
 DR EMBL; Z69757; CAA93625.1; JOINED.
 DR EMBL; Z69758; CAA93625.1; JOINED.
 DR EMBL; Z69759; CAA93625.1; JOINED.
 DR EMBL; Z69760; CAA93625.1; JOINED.
 DR EMBL; Z69761; CAA93625.1; JOINED.
 DR EMBL; Z69762; CAA93625.1; JOINED.
 DR EMBL; Z69763; CAA93625.1; JOINED.
 DR EMBL; Z69764; CAA93625.1; JOINED.
 DR EMBL; Z69765; CAA93625.1; JOINED.
 DR EMBL; Z69766; CAA93625.1; JOINED.
 DR EMBL; Z69767; CAA93625.1; JOINED.
 DR EMBL; Z69768; CAA93625.1; JOINED.
 DR EMBL; Z69769; CAA93625.1; JOINED.
 DR EMBL; Z69770; CAA93625.1; JOINED.
 DR EMBL; Z69771; CAA93625.1; JOINED.
 DR EMBL; Z69772; CAA93625.1; JOINED.
 DR EMBL; Z69773; CAA93625.1; JOINED.
 DR EMBL; Z69774; CAA93625.1; JOINED.
 DR EMBL; Z69775; CAA93625.1; JOINED.
 DR EMBL; Z69776; CAA93625.1; JOINED.
 DR EMBL; Z69777; CAA93625.1; JOINED.
 DR EMBL; Z69778; CAA93625.1; JOINED.
 DR EMBL; Z69779; CAA93625.1; JOINED.
 DR EMBL; Z69780; CAA93625.1; JOINED.
 DR EMBL; D14540; BAA03407.1; -;
 DR EMBL; L01986; AAA92511.1; -;
 DR EMBL; U04737; AAA18644.1; -;
 DR EMBL; S78570; AAB34770.1; -;
 DR EMBL; X83604; CAA58584.1; -;
 DR EMBL; S66432; AAB28545.1; -;
 DR EMBL; AP231988; AAG26332.2; ALT_TERM.
 DR TRANSFAC; T02337; -;
 DR MIM; 159555; -;
 DR InterPro: IPR001487; Bromodomain.
 DR InterPro: IPR003889; Fyrich.C.
 DR InterPro: IPR003888; Fyrich.N.
 DR InterPro: IPR001965; PHD.
 DR InterPro: IPR003616; PostSET.
 DR InterPro: IPR001214; SET.
 DR InterPro: IPR002857; znf-CXXC.
 DR Pfam; PF00628; PHD; 3.
 DR Pfam; PF00856; SET; 1.
 DR Pfam; PF02008; zf-CXXC; 1.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00542; FYRICH; 1.
 DR SMART; SM00541; FYRICH; 1.
 DR SMART; SM00249; PHD; 4.
 DR SMART; SM00508; PostSET; 1.
 DR SMART; SM00317; SET; 1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR PROSITE; PS50280; SET; 1.
 KW Proto-oncogene: Chromosomal translocation; DNA-binding; Bromodomain;
 KW Nuclear protein; Zinc-finger; Metal-binding; Transcription regulation;
 KW Alternative splicing.
 KW DOMAIN 17 102
 FT DNA_BIND 169 180 A.T HOOK (BY SIMILARITY).
 FT DNA_BIND 217 227 A.T HOOK (BY SIMILARITY).
 FT DNA_BIND 301 309 A.T HOOK (BY SIMILARITY).
 FT ZN_FING 1147 1194 CXXC-TYPE.
 FT ZN_FING 1431 1482 PHD-TYPE 1.
 FT ZN_FING 1484 1533 PHD-TYPE 2.
 FT ZN_FING 1566 1627 PHD-TYPE 3.
 FT DOMAIN 1703 1748 BROMODOMAIN (DIVERGENT).
 FT DOMAIN 3840 3969 SET.
 FT DOMAIN 1337 143 POLY-GLY.
 FT DOMAIN 561 564 POLY-PRO.
 FT DOMAIN 568 571 POLY-PRO.
 FT SITE 1362 1363 BREAKPOINT FOR TRANSLOCATION TO FORM MLL-

Query Match 4.8%; Score 353; DB 1; Length 3969;
 Best Local Similarity 20.2%; Pred. No. 2.6e-06;
 Matches 289; Conservative 155; Mismatches 466; Indels 522; Gaps 61;

```

QY 90 LLESDKHCPTCHQNDVSPDALIANKFLROAVNNFKNETGYTKRLKOLPPF----- 141
DB 360 IIPSSKRT-----DATIAKOLLORA-----KKGAQKLEKEAQAQGRKVKQ 402
QY 142 -----LFLVPPRPPLSQRNLQPRSRPILRQODPVY-----FRYTSPTCSDKT 186
DB 403 VKNIRQFIMPVVAISRIKTPRFIEDEDYDPPIKIAELESTPNRSRS-APSGSSSEK 461
QY 167 AGSCSD-----SGLSRLLPAPISISLTSSNQS-----LAPVSGNPSAPAPVDTATV 236
DB 462 SSAASQHSQMSDSSRSSSVDTSTDSQASEEIQVLPEERSDPEVHP-PLP-----I 515
QY 237 SISVHSEKSDGPFPRSDKLLPAAALTSEHSKASSIAITALEMEKGVGTSPMNSIFVG 296
DB 516 SOSPENESNDRRSRYS-----VSESPGSRP----- 542
QY 297 QSLHGLIPTTGPVIRINAAAPGGRPGMEHSNKLGYLVSPPOQIRRGESCTYSINCR 356
DB 543 -----TKKLSTLQSAPOQ----- 555
QY 357 HHSEKSOFTOSPLPATPCFVVPVPPPLYPPP-----HTLLPPLGVPV--PQESP 405
DB 556 -----QTSSSP-----PPPLLPPLLPASSISDHT--PWLMPPTTLPASP 595
QY 406 QPSSQPTAGYS--VPPPGF-----PPAPANISTAFCS-PGV--PTAHSTMTPT 451
DB 596 FLPASTAMOGKRKSIILREPTFRWTLKHSRSEPOYFSSAKAKGLIRKPIFDNRP 655
QY 452 QAPLISREFFREQNDKRESKFPYSSSYSSSTDSQGLAQIHIALTLSPSAHTLD 511
DB 656 LTP-----BDVGFASGFSASGTAASARLPSPLHSGRFTPMH--KRSP----- 695
QY 512 LLDHDPHPPEAEASAMIVHMDLMDIAHARSRPYRYSRSRSPREFGOSPTKEN 571
DB 656 LLARARFTPESEHSHRIFESVTLPSKRTSAGTSSGVSNNKRRKRYSPTRSEPRSPS--H 753
QY 572 VPREEKEREYFNRYREVPDYDIKAYGRSDV-----FRDPFEKERYREWEKRYR 621
DB 754 SMKTRSGRLSSSELSPLTPPSSVSSLSISVPLATSALEPFTPPS----- 800
QY 622 EMEYKTYGYANGAQRPP-----SANREDF--SPERLPLNIRNSPFTTGRREDYAAQS 674
DB 801 --HSLTQGESEAEKNQRPKQTSAPAEPPSSSPTPLPFWPTPGSQTERGRNKDKA---- 854
QY 675 HNRNMLGNYPEKLS--TRDSHNAKNPKSKESESENVPGDGKGNHKKRRKRENEKGE 733
DB 855 -----PELSKDRADAKSVKDKSRERDRER-----EKENKRESKKEKRRKKQSEIQ 900
QY 734 SESFLNP-----ELLETSRCKRGSSGIDETKTDTLFVLSRDPATPVDEPMDA 782
DB 901 SSSALYPGRVGRVKEKYGEDVATSSAKATGRKKS-----SHDSGT-----DI 945
QY 763 ESITKSVSDKRRKDKPKVKSDDTKRKSQSAATAKKUNVL--KTSKGQOEVDGDRS 841
DB 946 TSVTL-----GDTTAVKTKLILKGRGLMEKTIND--LG 977
QY 842 PSEPEPLKAKE--EATKIDSVKPPSSS-----OKDEVOTTPKRAHSKAK 886
DB 978 P-TASLSLEKTKLCISTPSSSYVKHSTSSIGSMLAODKLPTMDKRVASALLKAKAQOLCK 1036
QY 887 DTRROSQPTRSRKRTVPKTSQSQQPVFT--RPRSLRKINVLIAREKNEBEKRRKSV 944
DB 1037 IKRSKSLKQTDQ-----PRAQGESDSETSYGRPR--IKH-VCRRAAVLGRKRAVE 1086
QY 945 KPFESSSMKISKVEGEIYKPSPKRMEDVLEKLETP--EKDKLASSTTPAKKILANE 1002
DB 1087 PD-----DMPTLSALPWEEREKILIS-----MGND 1111

```

```

QY 1003 TGKRTGNAENASTKPESEKLES--TSKTIQEK--VKGAKRRVAGSBG-----SSSTLV 1054
DB 1112 DKSSIASGEDEAPLAPLPIKPIKPIVTRNKAPOPPYKAGRRSCQCGCCQVDECCGVT 1171
QY 1055 DYTSTSTGGSPVRKSEKTDKRTVITMEYNNNDTAPADVIIMIQVPSKWDKDF 1114
DB 1172 NCLDPRKFGGRNIRKQCK-----MRQCQMLQWMPSKAVYLQKAKAVKREKSK 1221
QY 1115 ESEEDVKTQPIQSVGRSSITKNVTTPKSPATAYTEKESQPEKLOKLPKASHLMO 1174
DB 1222 TSEKRD-----SKESSVYKVVV-----DSSQKp----- 1244
QY 1175 HBLRSKGSASSEKGRADREHSGSEKNDPKRKSQAOPDKSTYDRLSEOGHFTLSQS 1234
DB 1245 -----TPSAREDPARKSSSEPPPKPPEKSEBQNVNAPGE 1282
QY 1235 SKETRTSEKHESVKGSSNKF-----TPGRDKKVYDSDRYSSSKRRDERGELARRD 1287
DB 1283 SKQATTPASRKSSKQVSOQALVIPPQPTTGPPRK-----EVPKTPP 1324
QY 1288 SPPRGKESLSGKSKLREBRDLPKKGAEKSKSNSPPRDKKHHDKAPYETRRPCEETKP 1347
DB 1325 SEPKRKQ-----PPESGPEQSKQKVAAPRSITPVOK--PKREKRP-----PP 1367
QY 1348 VDK-----NSGKER-----EKHAEAARNGKSSG 1371
DB 1368 VKKQENAGTLNLTLSNGNSSKQKIPADGVHRIKRVDFEEDCEAEVWEMG 1419

```

```

RESULT 2
NCRR MOUSE
ID NCRR MOUSE STANDARD; PRt; 1453 AA.
AC P30415.
DT 01-APR-1993 (Rel. 25, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE NK-tumor recognition protein (Natural-killer cells cyclophilin-
  related protein) (NK-TR protein).
GN
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93133624; PubMed=8421688;
RA Anderson S.K.; Gallinger S.; Roder J.; Frey J.; Young H.A.;
RA Ortaldo J.R.;
RT "A cyclophilin-related protein involved in the function of natural
  killer cells ";
RT proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
RN [2]
RP REVISIONS TO C-TERMINUS.
RC STRAIN=BALE/C; TISSUE=Blood;
RA Anderson S.K.;
RL Submitted (0CT-1999) to the EMBL/Genbank/DBJ databases.
CC - FUNCTION: COMPONENT OF A PUTATIVE TUMOR-RECOGNITION COMPLEX.
CC INVOLVED IN THE FUNCTION OF NK CELLS.
CC - SIMILARITY: CONTAINS 1 CYCLOPHILIN-LIKE PLASE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; 104289; AAA37500.2; ALT_INIT.
DR HSBP; Q27450; 1A33.
DR MGD; MGI:97346; NKTR.
DR InterPro; IPR002130; CSA_PTIase.
DR Pfam; PF00160; pro_isomerase; 1.

```

DR PRINTS; PRO0153; CSAPPISMASE.
 DR PROSITE; PS00170; CSA_PPIASE.1; 1.
 DR PROSITE; PS0072; CSA_PPIASE.2; 1.
 KM Cyclosporin; Isomerase; Rotamase; Repeat; Transmembrane.
 FT DOMAIN 1 176 PPIASE; CYCLOPHILIN-TYPE.
 FT DOMAIN 222 241 ARG/LYS-RICH (BASIC).
 FT DOMAIN 422 459 ARG/LYS-RICH (BASIC).
 FT DOMAIN 964 1003 ARG/LYS-RICH (BASIC).
 FT DOMAIN 198 273 ARG/SER-RICH.
 FT DOMAIN 468 565 ARG/SER-RICH.
 FT DOMAIN 658 812 ARG/SER-RICH.
 FT DOMAIN 1303 1453 ARG-SER TANDEM REPEAT-RICH.
 SQ SPOUNCE 1453 AA; 163439 MW; DFL173FF814B283E CRC64;

Query Match 4.6%; Score 342; DB 1; Length 1453;
 Best Local Similarity 20.28; Pred. No. 2.6e-06;
 Matches 306; Conservative 119; Mismatches 573; Indels 458; Gaps 64;

65 KDIWDAVVIIPCCGNSCDECIRTTLESD---KHTCPTCHQ-----NDVSPD 109
 154 ENLKTDAASRYADYVIDGCVLATKLTDFEKKKKKPTCSGSDSSSSSSSSSE 213
 110 ALANKFLQAVNNKNETGYTKRLQPLPFLVPPRPLSQRLQ----- 158
 214 SEVERETIRRRHRRKPKVRAKKRREMS-----SSEPRKRRTVSPGSESDVNE 267
 159 -RMSPLRQODPVYFRYTSPTCSDTKTAGSCSDGTLSPAPIS-SITSQSSILAP 216
 268 KRSVDSNTKRRKRPVREELRPV-----PENRFLRDMRAITVEPQNIIPVAP 317
 217 PVSGNPSSAPRPVDITATVSVHS---EKSDGPRRDSOKLIP-----AA 260
 318 VVSQOKRSVSKSGKIKRGIRIHTPPRSKSHSEKDDDSLEPRMKEMQLRAYRP 377
 261 ALNTEHSGKASSIALITALEEKVGTPSPMNSIFVQ----- 297
 378 PSGRKMSKG-----DKLSDPCSWMDESLQSRSWSYNGYSDLTARHSDCH 427
 298 -----SILHGQILPTTGVR-----INAAIRPG 320
 428 HKHKRKKKKKKKKAKKQKHCHRRHROTKKRIYMPDLPEPSPHHRMKSCHERSRA 487
 321 GRPMEHNNKGLYVSPPOQ---IRGERSCYRSINRGHHSERSORTO----- 366
 488 SSSSHHSKRDMSKQDDOSASTHSSRDYRSKSHSRSSRSRAVSKSSSLN 547
 367 -----SPSLPATPCFVVVPPPLYP-PPPHLPPLPGVP-----PQGSFQFP--- 408
 548 RSKRSSSSRSRGPRTSISPKKPAQLSENKPVKTEPLRPVQNGNVLVQPAALNIPV 607
 409 -SSQPTTAGSVPPPGFPAPANISTACFSPGVTAHSNMPPTQAPLSSSEFFRQND 467
 608 LQSDPPPSRKM---PGQKP-----WKPSYRIQEMAKKTH--LLPVOSTYSLTNI 653
 468 KGRSEKPPY-----SGSSYRSSYTDSSQGLAHHALTLSPAHTLIDLHDPH 518
 654 KATVSSSYHKREKPSDGSAYSK--YSDSSG----- 685
 519 PPEAEARSAMIVMDLMDIAHARSRSPPRYRYSRSRSPPEFRGOSPTRKNVPREEKE 578
 686 -----SSGRSGS-----KSSRSRSSSRYSRSRSR-----LPSRSLSRSPSS 724
 579 REYF-NRYREVPPIYDIKAYY-----GRSYDFRDPFEK-----ERYREMRKRYREWEK 626
 725 RSHSPNKYSOSGHSRSSSTSVSSDGRAMFRSNRKKSVTSKRRHNSSEK--TLISK 782
 627 YKGYAAGAPRPSANREDESPERILPLNTNSPPTGRRE-DYA--AGOSH----- 675
 783 YVRG-----REKSS-----RHRKYSRSRLDVTSPSDGSHVQYVSAPE 821
 676 -----KNNNLGANTPE-----KLSTRDSHNAKNPKYSKESESENVPGDG 714

DB 822 KENQKVEALNDKQKGRGEGKPKPEWECPRSKKENSDH-SRDDSVSKKNCAGSWDS 880
 715 KGN-----KHKHKKRRNEKEGE-ESEFPLNELLTSSKRCRSGSIDETKDTTLVLP 767
 DB 881 ESNEDQVTSRKSDPRRGSEKEGEASSDSESVGSHIKAK-----PAKPTPTSTFLP 935
 768 SRDDATPVDEPMDAESITFKSVS-----DKDKREKDKPKYKSDTKRKSDGSAATAK 820
 DB 936 GSDQAMKSR-RPOSSASESSSCSNLNGEPEPKQKSHSDDLKGDHTKRAREKS-KAKK 993
 821 DNVLKPKGQEKVDGUREKSPRSEPLKAKAEATIDSKVPSSSSQKDEKVTGPR-K 879
 DB 994 D---KHKRAK-----RKAQFHWQPLFEGDDEEEMNG-----KQYQDPFEK 1034
 880 AHSKAKDTRRSQPRTRRSKRTVKTSSQSQPRTTRRPSLKLINYLAREKREKR 939
 DB 1035 RHSEKECAVKDGLP-----NVEKTCDEGSSSRKP-----KGTLEQD 1072
 940 KSYVDKFESSSMKISVEGETIYKPSPKRME---GDVEKLERPEKOKIAS--STTP 993
 DB 1073 PLAEQGHDPSSCAPLAVEDNTASSPSAQHLEHGGGEGDVLQTDNMEICTPDRTP 1132
 994 AKTKLNRFGKTIQANEMASTKEPSEKLESTSKIOEKVKAKARKYAGSGSS--- 1050
 DB 1133 AK-----GEVVSPLAN-----HRLDSEPVNTIIPDDECMAPRAGGEGESSMSE 1176
 1051 -----STVDYTSSTSGSPVKRSEKTDTKRTVTKTMEY-----NNDNTA 1093
 DB 1177 SKTIGESGVKODSSTVST--SPVETSGKKEGAEQSNLDMKRPLOGVGNLSVSTATIS 1234
 1094 PAEDVITIMIOVPOSQKMDKDFESEED-VKTTQDIQSVGRSSIIKNTYTKPSATANYT 1151
 DB 1235 SALDVKALSTYVEVKGPGRLIEKKNKVRGSLFDEVRKTARLNRPRNDESSDQTP 1294
 1152 -----EKSEQPEKIQKPKKESHELMQHELRSKGSASSEGRAKD---REHSGSEKDN 1204
 DB 1295 SRQDSQSRSPHR---SRKSETKSRHRTKRSVYSRSRSRSTSYRSRSTYSRCSR 1350
 1205 DKRSGAQPDKESTVDRLSEQGHFKTLSQSSKETRTSEKHESVGSSNNKFTTGRDKVD 1264
 DB 1351 DWYSRG-----KTRSSSYSGFSHRTSSRSRSSSD-----LHRSRSTY 1394
 1265 YDSRDYSSKRRDERGELARRKSDPPRGKESLSQOKSLREERDLPYKGAESKSSNSPP 1334
 DB 1395 YDSY-YRSRGRSR---SQRSDSYHRG-----RSYNRRSRSGRSGSDSES----- 1436
 1325 RDKKPHDKAPYETKR 1340
 DB 1437 -DRSYSHHRSPESSR 1451

RESULT 3
 NFH_HUMAN
 ID NFH_HUMAN STANDARD; PRT: 1020 AA.
 AC P12036;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neurofilament triplet H protein (200 kDa neurofilament protein)
 DE (Neurofilament heavy polypeptide) (NF-H).
 GN NFH OR NFH.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8328981; PubMed=3138108;
 RA Lees J.F., Sheldman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;
 RT "The structure and organization of the human heavy neurofilament
 subunit (NF-H) and the gene encoding it.";
 RL EMBO J. 7:1947-1955(1988).
 CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,

CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC NE-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC -1- PIV: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFLAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -1- PIV: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X15306; CAA33366.1; JOINED.
CC EMBL; X15307; CAA33366.1; JOINED.
CC EMBL; X15308; CAA33366.1; JOINED.
CC EMBL; X15309; CAA33366.1; JOINED.
CC PIR; S00979; QFHUH.
CC MIM; 162230; QFHUH.
CC InterPro; IPR001664; IF.
CC Pfam; PF00038; filament; 1.
CC PROSITE; PS00226; IF; 1.
CC Intermediate filament; Coiled coil; Neurone; Phosphorylation.
CC FT DOMAIN 1 100 HEAD.
CC FT DOMAIN 101 413 ROD.
CC FT DOMAIN 414 1020 TAIL.
CC FT DOMAIN 101 132 COIL 1A.
CC FT DOMAIN 133 145 LINKER 1.
CC FT DOMAIN 146 244 COIL 1B.
CC FT DOMAIN 245 266 LINKER 12.
CC FT DOMAIN 267 288 COIL 2A.
CC FT DOMAIN 289 292 LINKER 2.
CC FT DOMAIN 293 413 COIL 2B.
CC SEQUENCE 1020 AA; 111780 MW; 1177C9DCB3DCFD1D4 CRC64;

Query Match 4.6%; Score 337.5; DB 1; Length 1020;
Best Local Similarity 21.4%; Pred. No. 2.7e-06;
Matches 212; Conservative 164; Mismatches 458; Indels 155; Gaps 43;

QY 443 AHSNMPPTQAPLTSREERYRQNDKGRSKPPYSSSSSYRSTYDSSQGLAHQHTALT- 501
DB 120 AHNRSLEGAAL-----RQOQGRSA---MGLTERE--VREMGAVALRLCAANG 165
QY 502 -----LSPSAATHLDLHDHPHPPEEAARSAMIVMP-----DLMDIAHARS 545
DB 166 QARLEGEHLLLEIDIAHYRQLDDEARQREAEAAARALAFQAEAAARYDLQKKAALDE 225
QY 546 SPYRRYRSRSRSPPEFGQ-----SPTKRVNPREEKEREYFNRYREVPYDIKAVY 598
DB 226 ECGYLAR-RHHOEVGELLQIOGSGAAQOMQOETRDALCKDVTSALETRIAOLE----- 279
QY 599 GRSVDRDPFEKERYREMERKRYREMYEKYKAVAGQRPSPNRDPSPERLLPLNIRN 658
DB 280 GAHV--QSTLQSE--EW---FVRVLRDLRSEAKAVNTDMARSQOEITEYRQLQARTTE 331
QY 659 SPFTGRREDYVAGOSH---RNRNLGANTPEKLTSDS---HNAKDMPKSKSESENVPGD 713
DB 332 LEAKTKTSLERQSELEDRHADIASYQOALQDLAEIRNTKWMMAQLREYQDL---- 388
QY 714 GGNKHKRKRKRNRKEGESEFLNPELLETSRKCR--GSSGIDETKTDTLFLVLPGRDA 772
DB 389 -----LNVKALDIEIAVYAKILLE--GECRCIRGCPPIFSLPECLPIPSVSTH 435

QY 773 TPVRDEPMDAESTITFKSVSDK-----KREKDKPVKSDKTKRKSQDSATAKKDNVILKPSK 828
DB 436 IYKXSE---EKIKYKSEKETEIVVEQTEETQYTEETEE---EKPAKEEGKEEG 488
QY 829 GPOEKDGDRE--KSPRSE--PPLKA-----KEEA-TKIDSQKSSSQDKQKVTGPR 878
DB 489 GDEEAGEGEETKSPPAEASPEKEAKSPVKEAKSPAEPKPEAKSPAEPKSP 548
QY 879 KAKSKAKTRQOSQPRTRRSKRYVPKTSQSKQSPYTRRPSLKRINTLIREKNERK 938
DB 549 KAKSPKPEAK--SPEEAKSPKEEAKSPAEPKSPKAKSP--AKEEAKSPAEPKSP 603
QY 939 RKKSVDKDES-----SSMKIKVEGTEIVKP-----SPRKMGDEVKLETPPEKDKIS 989
DB 604 AKSPVKEEAKSPAEPKSPKEAKSPAEPKSPKEAKSPAEPKSPKEAKSPAEPKSP 662
QY 990 ---STTPAKKIKLNRGTGKIGNAENASTTKPESEKLESTSSKIKQKYGAKRRVAGS 1046
DB 663 PEKAKSPVKAEPKSPKAKSPVKAEPKSPKAKSPVKAEPKSPKAKSPVKAEPKSP 719
QY 1047 EGSSSTLVDTYSTSTSGSPVR--KSEKTDTKRIVITMEIYNNDNTPADVTIMIQ 1103
DB 720 ERAKSPVKEEAKTPPEKAKSPVKAEPKSPKAKSPKAKSPKAKSPKAKSPKAKSP 775
QY 1104 VPQSKDKDQDFSEEDVKTTO---PIQVGPSSIIKNVTTKPSATAKYTEKESEQPE 1159
DB 776 SPADKPEPAKSPVKEEVSPKAKSPKADAKAPE--KEIRKPEVKSVPVKEEKPQEV 833
QY 1160 KLQKLPEKASHIEMOHELRSKGSASSEKGRAKDRHSGSEKDNP--DKRRSGAOPDKE 1216
DB 834 KYKEPKKAKKEE-----KAPATPKTEKKKDSKEEAPKPEKAPKPEKPEKPEKPE 887
QY 1217 STYDRLSQGHKFTLSQSSKERTRTSEKHSYVSGSSKQDTPGDDKVVYDSDRYSSKRR 1276
DB 888 SKVEAKKEE-----AEDKKVPTPEKEAPKAVEVEDAKPKKTEVAKKEPDADAKAK-- 939
QY 1277 DERGELARRKDSPPRKGESLSQSKLREERDLPKKGAEKSSNSPPRD-KKPHDKAP 1335
DB 940 -EFSKAEKKEAPPEKDI--KEEAKKAPPEK--PKTEAKAKEDDKLSKSPKAKKAP 995
QY 1336 YETKRPCEETKPYDKNSGKEREKHAPEAR 1364
DB 996 KSSSTQKDSKPEKAT---EDKAKKAKG 1020

RESULT 4
MAPB_HUMAN
ID MAPB_HUMAN STANDARD; PRT; 2468 AA.
AC P46821;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain LC1].
GN MAP1B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retal brain;
RX MEDLINE=95104835; PubMed=7806212;
RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;
RT "Cloning of human microtubule-associated protein 1B and the
RT identification of a related gene on chromosome 15.";
RL Genomics 22:273-280(1994).
CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC STABILIZING MICROTUBULES.

CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -1- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -1- PWM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC from MAP1B by proteolytic processing. It is free to associate with
CC both MAP1A and MAP1B. It interacts with the amino-terminal region
CC of MAP1B (by similarity).
CC -1- SIMILARITY: TO MAP1A.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: I06237; AAA18904.1; -.
DR MIM: 157129; -.
DR InterPro: IPR000102; MAP1B_neuraxin.
DR Pfam: PF00414; MAP1B_neuraxin. 10.
DR PROSITE: PS00230; MAP1B_NEURAXIN. 6.
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.
FT REPEAT 1878 1894 MAP1B 1.
FT REPEAT 1895 1911 MAP1B 2.
FT REPEAT 1912 1928 MAP1B 3.
FT REPEAT 1929 1945 MAP1B 4.
FT REPEAT 1946 1962 MAP1B 5.
FT REPEAT 1963 1979 MAP1B 6.
FT REPEAT 1997 2013 MAP1B 7.
FT REPEAT 2014 2030 MAP1B 8.
FT REPEAT 2031 2047 MAP1B 9.
FT REPEAT 2048 2064 MAP1B 10.
FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT KKEE AND KKEI/V REPEATS).
SQ SEQUENCE 2468 AA; 270618 MW; 540839CBDF09D461 CRC64;

Query Match 4.58; Score 333.5; DB 1; Length 2468;
Best Local Similarity 22.18; Pred. No. 9e-06;
Matches 170; Conservative 127; Mismatches 282; Indels 191; Gaps 33;
QY 678 RNLEGNTP-----EKLSTRDSHNAKDNP---KSKESKES-ENYPGDKGKHKHHR 723
Db 519 KDLTGQPTPVYKQTKLKQRADSRSLKPAKRLPSKSVKRESKEERPEYTKVNHVERPP 578
QY 724 KRNEEKGESESEFLNPBELLETSSKRCGSSGIDETKT-DLFLVLPSSRDATPVADPEMDA 782
Db 579 KVESKEK-----VMVKKDKPVK-----TEKPSVTEKEVPSKEEPSVAEAEK 623
QY 783 ESTFKEVSODKREKDKPKYKSDTKRKSDGSAATAKKDNVLRKSGQKQEVDDREKSP 842
Db 624 QADIVKRAAKAEKTKVKEKPEKDKKEK---KPKKEVAKKDKPPRK-----EKKP 675
QY 843 RSEPLKKAKEEATKIDSVKPSSSOKDEKVTGTPRAKSHSAKDTROSQPTRRSKRT 902
Db 676 KKEEVKKEVKEIKKEKKEKPEKKEVKE---TPPEVKEVKEV-----KEEKE 720
QY 903 VPKTSSOSQDPVTRRRPSRLKIYLIAREKNEKREKRSKSDVKDFESSMKISVEGETE 962
Db 721 VKK-----EKEPEKKEIKKLPKAKKSSPPLSEAKRPAA 754
QY 963 VKPSKRMKGSDVEKLERTPEKOKIASSTPAKIKILNREGKIKGNAENASTHKEPSEK 1022
Db 755 LKPK-----VPKKEESVAKDSVAA-----KKEK 779
QY 1023 LESTSKIKOEKVGAKRKVAGSESSSTFLVDYTSSTGSGPVRSKSEKTDTRKRVYK 1082
Db 780 -----GKIYIKKKGKAENVAAGVGTGATL---AAYMAAGIAIGPAKLELEERSLMS 831

QY 1083 TMEEYND-NTAPADVYIMIQV--PQSKWDKDFESEEDVKTTOPIQS--VGKPSSTIK 1138
Db 832 SPEDLTDFEELKAEVDVTKIDKPOLELIED-----EKKKEIPEAVYIQKREVTK 886
QY 1139 NVTTKPSATYAKTYKESE--QPEKLOKLPKASHLMQHELRKSKSASSEK----- 1189
Db 887 GPASPEGLITTTGEGCEGTPEELPEVEKQGVADIEKFEDEAGGESESETDYERKA 946
QY 1190 ---RAKREHSGSE-----KDNP-----DKRKSQAPD--KESTYDR 1221
Db 947 ETEAEPEDEDEEHVCVSAKSHPTDEESAKAEADATYREKRESVAASGDRAEEDPDE 1006
QY 1222 LSEQGHFKTTSQSKETRTSEKHSYKSSNKDPTPGKDKKVDY----DSRDYSSKRRD 1277
Db 1007 AIEKGE---AEQSEDEADEDKADAR---EEYEPEKMEAEYVMAVVDKAAAGAE 1060
QY 1278 ERGLARRKDSPPR--GKESLSGQSKLRERDLPKKGAEKSSNSPP--RDKRPHDHKA 1334
Db 1061 QYGFLLT-----TPKQLAQSFGREPASSIHDETLP--GSESEATASDEENREDQPEEFTA 1115
QY 1335 ---PYETKRCPCEPTKPVDKNSGKEREKHAAEARNKSGSSGANCCHVYLTR 1380
Db 1116 TSGYGTQSTIEISSEPTPADEMS--TPROVMSDETNNETESPQSEFVNITK 1164
RESULT 5
ATRX_HUMAN STANDARD; PRT; 2492 AA.
ID ATRX_HUMAN
AC P46100; P51068; Q15886; Q9NTS3; Q9H0Z1;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE nuclear protein) (XNP) (Znf-HX).
CN ATRX OR RAD54L OR XH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND VARIANT S-1860,
RP AND VARIANTS ATR-X.
RX MEDLINE=97123494; PubMed=8966741;
RA Pickelits D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrelli O.W.J.,
RA Gibbons R.J.;
RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations
RT point to a common mechanism underlying the ATR-X syndrome.";
RL Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
RX MEDLINE=97386582; PubMed=9244431;
RA Villard L., Loebl A.-M., Cardoso C., Proud V., Chiaroni P.,
RA Colleaux L., Schwartz C., Fontes M.;
RT "determination of the genomic structure of the XNP/ATRX gene encoding
RT a potential zinc finger helicase.";
RL Genomics 43:149-155(1997).
RN [3]
RP SEQUENCE OF 860-2492 FROM N.A.
RX MEDLINE=95179111; PubMed=7874112;
RA Strayton C.L., Dabovic B., Gulisano M., Gecz J., Broccoli V.,
RA Giovannazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RA Bianchi M.E., Consalez G.G.;
RT "Cloning and characterization of a new human Xq13 gene, encoding a
RT putative helicase.";
RL Hum. Mol. Genet. 3:1957-1964(1994).
RN [4]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94214473; PubMed=8162050;
RA Gecz J., Pollard H., Consalez G., Villard L., Strayton C.L.,
RA Millaesau P., Khrestchatsky M., Fontes M.;
RT "Cloning and expression of the murine homologue of a putative human
RT X-linked nuclear protein gene closely linked to PKGI in Xq13.3.";
RL Hum. Mol. Genet. 3:39-44(1994).

RN [5]
 RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
 RX MEDLINE=95211835; PubMed=7697714;
 RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
 RT "Mutations in a putative global transcriptional regulator cause X-
 RL linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
 RL Cell 80:837-845(1995).
 RN [6]
 RP SEQUENCE OF 1375-2492 FROM N.A.
 RA Pearce A., Chapman J.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP E2H2 BINDING.
 RX MEDLINE=98167853; PubMed=9499421;
 RA Cardoso C., Timmit S., Villard L., Khrestchatskiy M., Fontes M.,
 RL Colleaux L.;
 RT "Specific interaction between the XNP/ATR-X gene product and the SET
 RL domain of the human E2H2 protein.";
 RL Hum. Mol. Genet. 7:679-684(1998).
 RN [8]
 RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
 RP HETEROCHROMATIN.
 RX MEDLINE=20040663; PubMed=10570185;
 RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
 RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
 RT "Localization of a putative transcriptional regulator (ATRX) at
 RT pericentromeric heterochromatin and the short arms of acrocentric
 RL chromosomes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
 RN [9]
 RP DISEASE.
 RX MEDLINE=20213147; PubMed=10751095;
 RA Villard L., Fontes M., Ades L.C., Geetz J.;
 RT "Identification of a mutation in the XNP/ATR-X gene in a family
 RL reported as Smith-Fineman-Myers syndrome.";
 RL Am. J. Med. Genet. 91:83-85(2000).
 RN [10]
 RP VARIANT ATR-X SER-1713.
 RX MEDLINE=97196774; PubMed=9043863;
 RA Villard L., Lacombe D., Fontes M.;
 RT "A point mutation in the XNP gene, associated with an ATR-X phenotype
 RL without alpha-thalassemia.";
 RL Eur. J. Hum. Genet. 4:316-320(1996).
 RN [11]
 RP VARIANT JM GLN-2131.
 RX MEDLINE=96224392; PubMed=8630485;
 RA Villard L., Geetz J., Mattel J.-F., Fontes M., Saugier-Verber P.,
 RA Munnich A., Lyonnet S.;
 RT "XNP mutation in a large family with Joubert-Marsidi syndrome.";
 RL Nat. Genet. 12:359-360(1996).
 RN [12]
 RP VARIANTS ATR-X.
 RX MEDLINE=97467722; PubMed=9326931;
 RA Gibbons R.J., Bachoo S.A., Picketts D.J., Aftimos S., Azenhauser B.,
 RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Kessler K., Kurosawa K.,
 RA Levin M.L., Masuno M., Nerl G., Plempont M.E., Slaney S.F.,
 RA Higgs D.R.;
 RT "Mutations in transcriptional regulator ATRX establish the functional
 RL significance of a PHD-like domain.";
 RL Nat. Genet. 17:146-148(1997).
 RN [13]
 RP VARIANT ATR-X LEU-246.
 RX MEDLINE=20123062; PubMed=10660327;
 RA Fichera M., Romano C., Castiglia L., Falla P., Ruberto C., Amata S.,
 RA Greco D., Cardoso C., Fontes M., Ragusa A.;
 RT "New mutations in XNP/ATR-X gene: a further contribution to
 RL genotype/phenotype relationship in ATR/X syndrome.";
 RL Hum. Mutat. 12:214-214(1998).
 RN [14]
 RP VARIANT SHS LYS-1742.
 RX MEDLINE=99347960; PubMed=10417298;
 RA Lossi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,

RA Prieto F., Fontes M., Martinez F.;
 RT "Mutation of the XNP/ATR-X gene in a family with severe mental
 RT retardation, spastic paraplegia and skewed pattern of X inactivation:
 RT demonstration that the mutation is involved in the inactivation
 RL bias.";
 RL Am. J. Hum. Genet. 65:558-562(1999).
 RN [15]
 RP VARIANT CMS THR-2050.
 RX MEDLINE=99326061; PubMed=10398237;
 RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,
 RA Curtis M.;
 RT "Carpenter-Waziri syndrome results from a mutation in XNP.";
 RL Am. J. Med. Genet. 85:249-251(1999).
 RN [16]
 RP VARIANTS ATR-X E-175; 178-V-K-198 DEL; S-190; P-219; L-246 AND C-249.
 RX MEDLINE=99219535; PubMed=10204841;
 RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Belongue J.,
 RA Lossi A.-M., Seaver L., Bonnefont J.-P., Romano C., Fichera M.,
 RA Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.;
 RT "Evaluation of a mutation screening strategy for sporadic cases of
 RT ATR-X syndrome.";
 RL J. Med. Genet. 36:183-186(1999).
 RN [17]
 RP VARIANTS ATR-X S-179; L-190; I-194; C-246; F-1552; S-1645 AND C-1847.
 RX MEDLINE=20451413; PubMed=10995512;
 RA Wada T., Kubota T., Fukushima Y., Saitoh S.;
 RT "Molecular genetic study of Japanese patients with X-linked
 RT alpha-thalassemia/mental retardation syndrome (ATR-X).";
 RL Am. J. Med. Genet. 94:242-248(2000).
 CC -I- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
 CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN
 CC BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.
 CC -I- SUBUNIT: PROBABLY BINDS E2H2. BINDS ANNEXIN V IN A CALCIUM AND
 CC PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSELINE-DEPENDENT MANNER (BY
 CC similarity).
 CC -I- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
 CC HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
 CC INTERACTING WITH HP1.
 CC -I- ALTERNATIVE PRODUCTS: 5 ISOFORMS: 1, 2, 3, 4 (SHOWN HERE) AND 5;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -I- TISSUE SPECIFICITY: UBIQUITOUS.
 CC -I- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-
 CC THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X
 CC SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPREHENSIVE SEVERE
 CC PSYCHOMOTOR RETARDATION, FACIAL DYSMORPHISM, UROGENITAL
 CC ABNORMALITIES, AND ALPHA-THALASSEMIA. AN ESSENTIAL PHENOTYPIC
 CC TRAIT ARE HEMOGLOBIN H ERYTHROCYTE INCLUSIONS.
 CC -I- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF CARPENTER-WAZIRI
 CC SYNDROME (CMS), AN X-LINKED RECESSIVE CONDITION CHARACTERIZED BY
 CC MODERATE MENTAL RETARDATION, SHORT STATURE, BRACHYDACTYL WITH
 CC EXCESSIVE SKIN CREASES, AND WIDENING OF THE KNUCKLES.
 CC -I- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JOUBERT-MARSIDI SYNDROME
 CC (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY
 CC SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS,
 CC MICROCENTRALISM AND EARLY DEATH.
 CC -I- DISEASE: DEFECTS IN ATRX ARE A CAUSE OF SMITH-FINEMAN-MYERS
 CC SYNDROME (SFM). CLINICAL FEATURES INCLUDE SEVERE MENTAL
 CC RETARDATION, MICROCEPHALY, GROWTH FAILURE, FACIAL ANOMALIES AND
 CC BILATERAL CRYPTORCHIDISM. DUE TO THE CLINICAL OVERLAP WITH ATR-X
 CC SYNDROME, SOME PATIENTS ORIGINALLY DIAGNOSED AS HAVING SFM, MIGHT
 CC BE AFFECTED BY A VARIANT OF ATR-X SYNDROME WHICH LACK HEMOGLOBIN H
 CC INCLUSIONS.
 CC -I- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF SUTHERLAND-HAAN X-LINKED
 CC MENTAL RETARDATION SYNDROME (SHS). IT IS CHARACTERIZED BY SEVERE
 CC MENTAL RETARDATION WITH SPASTIC PARAPLEGIA, MICROCEPHALY, SHORT
 CC STATURE AND CRYPTORCHIDISM.
 CC -I- SIMILARITY: BELONGS TO THE SNE2/RAD54 HELICASE FAMILY.
 CC -I- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: U72937; AAB49970.2; -
 DR EMBL: U72938; AAB49971.2; -
 DR EMBL: U72935; AAB40698.1; -
 DR EMBL: U72904; AAB40698.1; JOINED.

Query Match 4.4%; Score 326; DB 1; Length 2492;
 Best Local Similarity 19.0%; Pred. No. 1.7e-05;
 Matches 287; Conservative 227; Mismatches 566; Indels 434; Gaps 63;

62 LICKDITDVIIPCCGSSCDECI-----RTTLESDEKHTCPTCHQNDVSPDALI 112
 229 LICODF-----CHNACCKCICILNLGRKELSTIMDNQMYCICH-----PEPL 274
 113 ANKFLQAVNN-EKNEGYTKRLKQLPPLFLVPPRPISQRLQPRSRSPILROODPV 171
 275 D---LVYACNSVEN-----LEQLQKKKIKIVDSEKSNK 307
 172 VFRYT-----VSPYCS--DTKTAGSCSDGTLR---LP-----APSSISLTJNOS 212
 308 VYEHSTSFSPKRTSSNCGEKLDSGSGSVITYSALVPEKEMIKAKKALLETIANMN 367
 213 S-----LAPPVSGNPSSAPAPVDITATVSIASHSEKS-----DGPR--DSDNKL 257
 368 SSYVFKLQATDNSEISATKRLQKAFKSVLADIKAKHLALEDLNDSEPRADAVNK-- 425
 258 PAALITSEHKGASSTIYTLAMEKGVPTSP-----WNSITFGOSLILHQ 303
 426 -EKNTKEH------VIDAKFETKARKGEPALKEKDKISEAKLSRQVSEHMH-Q 476
 304 LIPTGVRINARPGGGR-----PGWESHNKLYL---VSPPOI----- 341
 477 NVPTL-EQRNKSTGGHKKSDRKEEPQYRPANTSEDLMDIYVPSVSEDTFENLETA 535
 342 -----RGRERS-----CYRSINGRHNSRSQR-TQSPSLPATPCFV 377
 536 MEYOSSVDHGGDSSGTEQEVESVYKLTISKNRGIKSKTAKYTKELYKLP--V 593
 378 PVPPRPILYPPPTLPLP-----GYPP----- 401
 594 SLFNSPTKAGDCQEVDPQDKGYSGLNPKLEKGLQENSDEHLENVSLLEESDL 653
 402 QFSPQF---SSQPPFAGYVPPPGFPPAPANISTACFSGVPTAHSTMTQAPILSR 458
 654 RBSPRVKTTLRKRTETN-----PVTNSDDEC-NETVAKKOLSLPV-----R 696
 459 EEFYREONDKG---RESKPPYSGSSYSRSSYTDSSQGLAQHIALTLSPSAHTLIDLH 514
 697 KKDKNSSDSALDNPKRNKLPKSKQSETVDQNSDEMLALILGVSMSHSSSDTDINE 756
 515 DHHPPPEAARASAMVHMDLMDIAHARSRPYRKRYSRSPPEF---RGOSPKNRN 571
 757 IH-----TNHKTLDLKTQAKDKGKRRKRSSTSGSFDYKKGSAKSSI 802
 572 VPREEKREYFNRYREVPPYDIKAYGVSVDPPDPREKRYRMEKRYEMWEKYYKGY 631
 803 ISKKKQTO-----SESNYDSELEKE----- 824
 632 AVGAOPRANRDEFSERLLPLINRNSPTGRREDYAGOSHNRNLGNYPEKLSTR 691
 825 -IKSMKIGAR---TJKRILP-----NTK 845
 692 DSNANADNPKSKESKESENVGDKGNKHKHRRRNEEKGESSESLINPLETSTRCRG 751
 846 DFSSSEDEKHSK-----KGMDOGHKMLKTSQESSSDAEKROK--RETFSSAEG 893
 752 SSIDETKDTLFLVPSRDATPVRD--EPMDAESTFKVSDKDKRE-KDKKRVASDKT 808
 894 TVDKDTTILMLRDLPLPKQOASASTDGVDLGSKSEOSTSLEVRKVAETREKSKHLTKT 953

QY 809 KRKSDGSATAKDNVILKPSKQPEKVDGDRKSPRESEPLTKAKAEATKIDSKVP---SSS 866
 954 CKRVQGLSDIAKFLPKDQSDSTESD-DKKQKKGEEKKKSDFPKKVTIKMGOYESS 1012
 QY 867 SQKDEKVTGTPRAH--SKSKADTRROSQPTRRSKRVPTSSQKSOQPVTRRRLRK 925
 1013 SDGTEKLPEREELICHPEKGIQKQIKNGTJTDKSKSKIRDTSKKDE-----L 1060
 QY 926 NYLIAREKNERKRSKVDPESSSMKISKVEGTETVRSPPRK--MEGDVKELETPRE 983
 1061 SDYAEKSTGDCDSDSEDKSKNGAIGREKKCKLIGKSSRRKQDCSSDTEKY--SMK 1118
 QY 984 KDIASSTYPAKRIKL-----NRETGKIGNAENASTTKESPE---KLESTSKIKQ 1032
 1119 EDGNSDKRLKRIELIERNNLSKSRNTEIQQSSSSDAEESSEDKKKKQRTSSKKA 1178
 QY 1033 EKYKGR-----AKRVAGSESSSTLVDTYSTSTGSGSPARKSEKIDTKRYVIKTM 1084
 1179 VIVKEKRNLSRTSTRKQDITSSSSDIEDDQNSIGGSSDEOKIKPVTENLVLSH 1238
 QY 1085 EEFYNDTAPAEVYIIMIQVQSKMDK-----DESEEDVKTTOPIQ 1128
 1239 TGFCQSSGDALSKSVYVYDDDDDDNDPENRIAKKMLLEIKANLSSDSDGSDDEPEE 1298
 QY 1129 SVGKPSIINKVTTKPSATAK-YTEKESQPEKLOKLPKASHLMOHELRSKGSASS 1187
 1299 --GKKRTGKQNEENPGDEEAKNOVNSESDSDSESKKPR-YRRHLLRHKLTVSDGEGEE 1355
 QY 1188 KGRKAKREHSGSEKDNDRKKSQAQPDKESTVRLSKOGHFKLTSOSK---TPTSEK 1243
 1356 K-RTKPEH--KEVKGRRKRVSESESD--FQSGVSEEVSEEDQRPRTSNAK 1409
 QY 1244 HESVRCSSNKDFTPGRDKKDYDSRDYSSSKRRDERGELARRDSPRGKESLQSKSL 1303
 1410 AE-----LENGRSY--KOKKKRRIRYQDSSSENNKSNSEEBEEK 1449
 QY 1304 REERDLPKGAESK---SNSSPPRDKP---HDKAPYETKRPCEETKRVYDNKSG 1354
 1450 EEEEEEEEEEEEDENDDSKSPGKRRKIRKLKDKLRTETQNALKEE-----E 1501
 QY 1355 EREKHAARNGKE 1368
 1502 ERRKRIRERERE 1515
 Db
 RESULT 6
 MST2_DROHY
 ID MST2_DROHY STANDARD; PRT; 1391 AA.
 AC 008696;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Axoneme-associated protein mst101(2).
 GN MST101(2).
 OS Drosophila hydei (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7224;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=95045538; PubMed=7957199;
 RA Neesen J., Padmanabhan S., Buemann H.;
 RT "Randomly arranged repeats of a novel highly charged 16-amino-acid
 RT motif representing the major component of the sperm-tail-specific
 RT axoneme-associated protein family Dmst101 form extended
 RT alpha-helical rods within the extremely elongated spermatozoa of
 RT Drosophila hydei";
 RL Eur. J. Biochem. 225:1089-1095(1994).
 CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY
 CC SPERMATIDS.

CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
 CC -1- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT
 CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM
 CC REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X73481; CA51876.1; -.
 DR PIR: S34154; S34154.
 DR HSSP: P01032; ICSA.
 DR Flybase: FBgn0020733; Dhyd\mst101(2).
 KW Sperm; Repeat; Multigene family: Polymorphism.
 FT DOMAIN 332 1268 [KR]-K-X-C-X-X-X-K-X-X-X-X-X-X-E.
 FT SEQUENCE 1391 AA; 159000 MW; 1B2A368F30F48678 CRC64;

Query Match 4.4%; Score 321; DB 1; Length 1391;
 Best Local Similarity 22.5%; Pred. No. 1.5e-05;
 Matches 162; Conservative 123; Mismatches 326; Indels 108; Gaps 28;

OY 720 KKKRRRRNKEGE-ESESLNPELETSRCKGSSGIDETK--TDTLFVLSRDPDAPV 775
 DB 285 KKKKKKKKEKEKELELEILKEQAEKEAKINGV--VKEVKKCKCKEALKKCKDGRKM 342
 OY 776 RDEPDASITFKSVSDKRRKDKPKVSDTKRSDSAKAKNNVLPSPGPEXVD 835
 DB 343 KEE--AEKKCAALAKKCKK-EDEKCKACKELAKKKKEDEKKCEEAANKKEKAAK 398
 OY 836 GPREKSPSEPLKAKKEATKIDSVKSSSQDEKVTGTPRK-AHSKAKDTRRSOP 894
 DB 399 CEKAKERKEAEKKCKCEAAKKE--KEAERKCKEELAKNNKKAKEKKCKEAAKKE 456
 OY 895 RTRRSK--RTVPKTSQKSOQPVNTRPRSL-----RKINYLIA 930
 DB 457 AAEKKCEELAKKIKKAAEKCEETAKKCKEVAERKCKEELAKKIKKAAEKCKKAK 516
 OY 931 RKKNEREKRR--KSYDKDEPSSSMKISKVEGTETVPSRKMKEDGVELETRPEKDIA 988
 DB 517 KKEETAEKKCKEAAKRRKEAEKK--KCE-----KAAKRRKEAAKCKCEKSAKRRKA 569
 OY 989 SSTPAKK-IKLNRETGKIGNAENASTTKEPSE--KLESTSSKIR---QEKVKAKAR 1041
 DB 570 AAKKCKEAKERKEAEKKCKCEAAKKEVAERKCKEELAKKIKKAAEKCKCKEAAK 629
 OY 1042 KYAGSE---GSSSTLVDTSTSSGSPVRKSEKTDTK--RTVIKTMEEYNNNDTAP 1094
 DB 630 EKEAEERKCGELAKKIKKAAEKCKCKLAKKEKTEAKCKEAKKCKEAAEKCKCAE 689
 OY 1095 AEDVITIMIVQPSKMDKDFESE--BEDVKTQPIQSVKPSI---IKNVTTKPRATK 1149
 DB 690 A-----AKKEKEAEKKCKCEAAKKEEAERKCKEELAKKIKKAAEKCKCKKL 738
 OY 1150 YTEKESEDEPEKLOKLPKEASHELMQ---HELRSKGSASSEKGRKADSHSGSEKNDP 1205
 DB 739 AKKKKAGEKNKTLKNGKKKGLKKEKKCKRELAKKKAKEKKCKEAAKEKEKCKE 798
 OY 1206 ---KRSQAQDPK-ESTYDRLESGHFTLSOSKSETTSEKHESVROSSKNDFTPGRD 1260
 DB 799 KTAKKKEAEKCKCKEKTAKKRRKEAEKCKCEAAKRRKEEAKEKCKEKTAKKRETAEK 858
 OY 1261 KKVDYDSRDYSSSKRRDEGELARRKDSPPRGKESLSGKSKLREPRDLPKKGAESKSN 1320
 DB 859 KKEE-----KAKKRKOAAEKCKCKEAKAKKRE--AAEKKCAEAAKEKELEAEKKCKE 910
 OY 1321 SSPPRDKKPHDKAPYETKRPCEE--TKPYDKNSGK-----EREKHAAEARNKSESSG 1371
 FT SEQUENCE 1391 AA; 159000 MW; 1B2A368F30F48678 CRC64;

DB 911 EAKKEKEVAERK-----KCEELAKKIKKAAEKCKCKLAKKEKAGEKNKLLKAG 962
 RESULT 7
 MAPB_MOUSE STANDARD; PRT; 2464 AA.
 AC P14873;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Microtubule-associated protein 1B (MAP1B) (MAP1.2) (MAP1(X))
 DE [Contains: MAP1 light chain LC1].
 GN MAP1B OR MTAP1B OR MTAP5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND DOMAIN.
 RC STRAIN=SWISS WEBSTER; TISSUE=Brain;
 RX MEDLINE=90094539; PubMed=2480963;
 RA Noble M., Lewis S.A., Cowan N.J.;
 RT "The microtubule binding domain of microtubule-associated protein
 RT MAP1B contains a repeated sequence motif unrelated to that of MAP2
 RT and tau."
 RL J. Cell Biol. 109:3367-3376(1989).
 CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
 CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
 CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
 CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
 CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
 CC STABILIZING MICROTUBULES.
 CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
 CC WITH MAP1A AND MAP1B PROTEINS.
 CC -1- DOMAIN: Has a highly basic region with many copies of the sequence
 CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
 CC responsible for the binding of MAP1B to microtubules.
 CC -1- PTM: LC1 IS COEXPRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
 CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
 CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
 CC OF MAP1B.
 CC -1- SIMILARITY: TO MAP1A.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X51396; CA35761.1; -.
 DR PIR: S07549; QRMSP1.
 DR MGD: MGI:1306778; Mtap1b.
 DR InterPro: IPR000102; MAP1B_neuraxin.
 DR Pfam: PF00414; MAP1B_neuraxin; 10.
 DR PROSITE: PS00230; MAP1B_NEURAXIN; 7.
 KW Microtubules; Repeat; Phosphorylation.
 FT CHAIN ? 2464
 FT REPEAT 1874 1890 MAP1B 1.
 FT REPEAT 1891 1907 MAP1B 2.
 FT REPEAT 1908 1924 MAP1B 3.
 FT REPEAT 1925 1941 MAP1B 4.
 FT REPEAT 1942 1958 MAP1B 5.
 FT REPEAT 1959 1975 MAP1B 6.
 FT REPEAT 1993 2009 MAP1B 7.
 FT REPEAT 2010 2026 MAP1B 8.
 FT REPEAT 2027 2043 MAP1B 9.
 FT REPEAT 2044 2060 MAP1B 10.
 FT DOMAIN 589 787
 FT SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFBDBA7 CRC64;
 KKEE AND KKEI/V REPEATS).
 LYS-RICH (HIGHLY BASIC, CONTAINS MANY


```
Db 525 SOSYSGSSSRRTASSSSHSRGRSKSRSSSGHRRKRAKSPRKTASOLSEKKPYKTEP 584
QY 384 L-----YPPPHPLLP--GVPPQ-----ISPQPPSSQ-----411
Db 585 LRATMAQNENNVVQPVVAENIPVPLSDSPPPSRMRPGOKPMPSYERIOEMAKTTHTLL 644
QY 412 PPTAGYVPPPGPPAPANISTACFSPGVPPLANSNTMPTQALLSREERYRQONKGRE 471
Db 645 PLOSTYSL-----ANIK-----ETGSSSYHKKREKNSDQSTYSKSDRSSSE 667
QY 472 SKFPYSGSSYSRS-----SYTDSQGLAQHIALTLSPSA-AHTLDLHDHPH-----518
Db 688 SS-PRSRSRSSRSRSRSRYT-RSRRLASS-HRSRSRSPSRSSRNKYSIDHSCSSRSY 744
QY 519 ---PPEBARASAMIVHMDLMDIAH-----ARSRPPRYRRSRSSRS 558
Db 745 TSLSSDGRARRKRLRSKSKNSVSHKSHSSSEKTLHSKYVKGRRSSGCVFRY-SESR 803
QY 559 PPEFRG---QSPTRKRVNPREKEKEVEFNRYREVPPYDIKAYYGRGVDFPDEKEERYE 615
Db 804 SLDISSDSESSVQATQSAQEKQ-----GQMERTHNKKQENKNGEE 845
QY 616 WERYKREMYEKYKGYVGAQPRPSANREDSPERLLPLN---INSPFTGRREDYAA 671
Db 846 KSKSEKREC-----PHSKKRLKENTLSDHLKNG-SKPKRKNYAG 882
QY 672 ---GQSHRNRLGNYPEKLTROSHNAKDNPKSEKESENVPGDGKGNKHKRRRN 727
Db 883 SKWDSSESNBERDVTKN-----SKNDSH-PSSDKEGEAT-----915
QY 728 EEKGESEFLLNPELTETSRKRCGSGIDETKTDLFLVPSRDA---PPVRDEPMDAES 784
Db 916 ----SSSES---EVEEIHAKVPT-----TKSSTNTSLPDDNGAKSSQKRTSTDSSE 962
QY 785 IFFKVSVDKDKREK---DKPKVSDKTKRRKSDGATAKKDNVLPKSPQOEKVDGDKREK 840
Db 963 SCNSNSNNGKPKQKHGSKENLKRHTTKVKE-KLKGKD---KHKHAKV-----RKQ 1012
QY 841 SPRSEPLKAKKEAKIKDSVKRSSSSQKDEKTTGTPRAHNSAAD---TRQSQPRT 896
Db 1013 AFHWQPPLEEGEEBEEDID-KOVTOESKEKVS---ENNETIDNLIKTKSSSEDL 1066
QY 897 RRSKRTVPKTSQSQOPVTRRRSRSLKINYLAREKNEREKRRKVDKFESSMKISK 956
Db 1067 SGKHDVITYSSD-----LDQFTKDDSKLSISPALMTVEENVAC 1104
QY 957 VEGTEIVKQSPKRMEDVEKLETERPEKDIASSTPAKKIKLRNTEGKKIGNA-----1010
Db 1105 LQNIQHYEES---VPNGVEDVLQTDNMEIC---TPDRSSPAKVEETSPFGNARLDTPD 1157
QY 1011 ----ENASTTKPEPEKLESTSKIOEKYKGAKKRVAGSESSSTLVDTYSTSTGGS 1065
Db 1158 INIVLQDMATEHPQAEVVKQESSMSSEKVLGE---VGKODSSS---ASLASGES 1207
QY 1066 PVKSESEKIDTKRTVITKTEMEYN-----NDNTAPAEVYIIMIOVPOSKMOKDFESE 1117
Db 1208 TGKKEVAEKSQILIDKMKKPLQGVGNLAPNMAATSSAIVEKVLITYPEAK-----PQGL 1262
QY 1118 EEDVKTQPIQSVGKPSSTIKVNTTKPSATAKTEKESQPEKLOKLPKASHLMOHEL 1177
Db 1263 RLEIKSKKNV---RPSGLFDEVRKATRLNRBRNQESSSDQETPSRDDSS-----QS 1311
QY 1178 RSSGASASSKGAOKREHSGSEKNDPKRKSQAQPDKESTVRLSEQGHF---KTLISOS 1234
Db 1312 RSPSRSRSSSEKSRHRTKRVSTSHSRSRSSSTSSYSRSRSRGMYSNGKTRRSR 1371
QY 1235 S-----KETRTSEKHSVVGSSNKKDFTT-GRDKKVDYDSSSKRRDRGELARRKDS 1288
Db 1372 SSVRSYKSHHTSSRSRS-RSSS---YDPHRSRSYTYDST-YRSRSRSR-----SQRSDS 1422
QY 1289 PRPGKSSLGQSKLREERDLPKKGABSKSSNSPPRDKKPHDHAKAPYETKR 1340
Db 1423 YHRG-----RSYNNRRSRSCRSYGSDES-----DRSYSHHRSSESSR 1460
```

```
RESULT 9
ID MLH_TETTH STANDARD; PRT; 633 AA.
AC P40631;
DT 01-FEB-1995 (rel. 31, Last created)
DT 01-FEB-1995 (rel. 31, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Micronuclear linker histone polypeptide (MIC LH) [contains:
DE Micronuclear linker histone-alpha; Micronuclear linker histone-beta;
DE Micronuclear linker histone-delta; Micronuclear linker histone-gamma].
GN MLH.
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymena; Tetrahymena.
OX NCBI_Taxid=5911;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=CU401;
RX MEDLINE=94088506; PubMed=8264578;
RA Wu M., Allis C.D., Sweet M.T., Cook R.G., Thatcher T.H.,
RA Gorovsky M.A.;
RT "Four distinct and unusual linker proteins in a mitotically dividing
RT nucleus are derived from a 71-kilodalton polypeptide, lack p34cdc2
RT sites, and contain protein kinase A sites.";
RL Mol. Cell. Biol. 14:10-20(1994).
CC -1- SUBCELLULAR LOCATION: NUCLEAR; MICRONUCLEI.
CC -1- PTM: ALL FOUR HISTONES ARE PROCESSED FROM THE PRECURSOR MOLECULE.
CC THEY ARE PHOSPHORYLATED IN GROWING AND DIVIDING CELLS BUT NOT IN
CC NONROWING (STARVED) CELLS. THE N-TERMINAL OF ALPHA AND DELTA IS
CC BLOCKED.
CC -1- SIMILARITY: CONTAINS 2 HMG BOXES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M87306; AAC18874.1; -.
DR InterPro: IPR000910; HMG_12_box.
DR Pfam: PF00505; HMG_box; 1.
DR SMART: SM00398; HMG; 2.
KW Nuclear protein; Chromosomal protein; DNA-binding; Repeat;
KW Phosphorylation.
FT CHAIN 1..399 MICRONUCLEAR LINKER HISTONE-ALPHA.
FT FT 1..199 MICRONUCLEAR LINKER HISTONE-DELTA.
FT CHAIN 200..399 MICRONUCLEAR LINKER HISTONE-GAMMA.
FT FT 400..633 MICRONUCLEAR LINKER HISTONE-BETA.
FT DNA_BIND 12..74 HMG BOX 1.
FT FT DNA_BIND 96..164 HMG BOX 2.
FT DOMAIN 181..391 LYS/SER-RICH.
SQ SEQUENCE 633 AA; 70624 MW; F3B44B/AAE7B3FC CRC64;

Query Match 4.3%; Score 315; DB 1; Length 633;
Best local Similarity 21.6%; Pred. No. 1,2e-05;
Matches 167; Conservative 126; Mismatches 283; Indels 196; Gaps 28;

QY 650 RLPLPLNR--NSPFTGRGRDYAAGSHRNKRLNGATPEKLTARD---SHNAK-----697
Db 4 RRTKLNIQPRKPPS---NTYQAFVLEKKNALGKNDKRYQADYVADNKLNNKNERLQKL 59
QY 698 -DNPKSKKE-----SENVPGDGKGNKHKRRKRNDEKEESESFLNPELLTSTRKCG 751
Db 60 VDNAEKYKELFHYNNHIOGKR-KQYIVQVAVPEPKPPISSF---RFLDENQKTA 115
QY 752 SSGIDETKTDTLFLPSRDDATPVPRDEPMDAES-----ITFKVSQDKRREKDKPK 802
Db 116 AKKRDLTNAKILKMSBDFNNLPQKEVKYVEDAYQKEVYQAEYLVEFKMNEKYGQAQAKQ 175
```



```

Db 425 RSRYS-----MSERSEGSRA-----440
QY 308 TGPRIINARGGGRPGMEHSHNKLGYLVSPQQLRGRGRCYRSINGRHSHSESOQTOS 367
Db 441 -----TRKLPLOSAPO-----OKTSS 457
QY 368 PSLPATCFVVPVPPPLYPPPPHTLPLPGVPPPOFSQPFSSQPPVAGVPPPGFPYA 427
Db 458 S-----PPPLPTFP-----PL-----QPASGIDHPWMLMP- 485
QY 428 PANISTACFSFVPPTANSNTPTQAPLISREEFYREONDGKRESKPEYSSGYSRS--S 485
Db 486 -----PTIPLA-SPFLPASAAPWQGRKSL-----REPTFRWTSLKHSRSPQ 528
QY 486 YTDSOGIAGIITHLTSPSAHLLDLHHPPEE-----AASAMIVH 532
Db 529 YFSAAKAKE-----GLIRKP-----IFDNFRPPLPEDGFGASGASCTAASARLFSPH 581
QY 533 MPDLMDIAHARS-----RSPPYRRYRSR-----SPPEFGQSPTKRNVPEEKREYFN 583
Db 582 SGTFRDI-HKRSPLLRAPRTFPAHSHKIFESVTLPSKRTISGASGSSGVNKRKRKYFS 640
QY 584 RYREVP--PPYDIKAYGR-----SVDFRDPF 608
Db 641 PIREPSPSPSHMRTSGRLSTSELSPLTPSPSSVSLIPVPLASALNPTTFPSSH 700
QY 609 EKEVYRMEKRYRMEYKYGYAVGAQPPRANREDFSP--RLPLTINNSPTTRRR 666
Db 701 LTQSGDSTEKQR-----ARKQTSAPAEPPSSNSPALFPWFTPGSQTEKGRK 747
QY 667 EDYAGOSHRNRNIGVNYPEKLS--TRDSHNAKDNPKEKSEENVPGDGKHKHKKHRK 725
Db 748 KDTA-----PELSKDRADKVEKDKSERRER-----EKNKRESREK 789
QY 726 RNEKGESESEFLNP-----ELLETSKRCGSSGIDETKTDTLFVLPSPDQATP 774
Db 790 RKKSQSDIQSSSALPYGVSKYKAVAGEVGTSSAKKATGRKSSS-----LDGADVAP 844
QY 775 VRDPMAEISTTFKSVSDKDKREKDKPKVSKDKTKRKSQATAKKQNVL-KPSKGPEK 833
Db 845 V-----TLGPTTAVAKILLIKKGGNLEK 868
QY 834 VDGD-----REKSP-RSEPLKKAKEATKIDVSPSSS--QKDEVTGTPRKAH 881
Db 869 NNLDLGRAPELEKERPTCLSAPESSSYKHSSTISGSLAODKLPMDKRYASILKAK 928
QY 882 SKSAKDRROSOPTRRSKRTVPPTSQKSQPVRT--RRPR--SLRKINYLAREK--- 933
Db 929 AOLCKIEKSKLKQTDQ-----PRAQGESEDSSETSVGRPRIKHYRAAVALGKRRAVF 983
QY 934 -----NREKRRKSVDKDFESSSMKISKVEGTELVKP-SP-----KR 969
Db 984 PDDMTLSALPWEREREKILSSMGNDKSS-----VAGSEDAEPLAPIKIPYTRKA 1037
QY 970 KMEGDVEKLERT-----PEKDKIASSTP----- 993
Db 1038 POEPPVKKGRKSRRCGCGPCQVPEDCITCNLDKPFGRGRIKCKKCKRKNCONLQW 1097
QY 994 AKKILNRTGKIKGNAENASTKEPSEKLESTSKIKOEKVGKA-----KRVAGS 1046
Db 1098 PSKASLKQT-KAVKKEKKKSKTTEKKESKSTSVKSPLEPAQAAPPRPREPAKSSS 1156
QY 1047 EGSSSTLVDTSTSTSGSPV-----RKSEKTDTKRTVITKMEBYNDN 1091
Db 1157 EPPRRKPEVE--EKSEEGGAPAPAPAPPEKQVSAVASRKSQVQAPAVVPPQ-----PS 1210
QY 1092 TAPAEVYIMQVPOSKMDKDFESEEDVYTPQIQSVKPSSTIKIVYTKPSTATKYT 1151
Db 1211 TAPQKK-----EAPKA-----VPSEPKKQPPPEPGEQSKQKVAADLPSTPVKQK 1257
QY 1152 EKESOEPEKLOKLPKEASHELMOHELRRSSKSSASEK 1188
Db 1258 PKDKEKPPPVSKQENAGTLNLTN---PLSNIGSSKQK 1291

```

```

RESULT 11
ZAP3_HUMAN
ID ZAP3_HUMAN STANDARD; PRT; 1822 AA.
AC P49750; P49752; Q9P1V7;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear protein ZAP3 (ZAP13).
GN ZAP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Madan A., Qin S., Abbasi N., Baradaran L., Birditt B.,
RA Bloom S., Dors M., Dichtoff R., Fleetwood P., Harrison G., James R.,
RA Kaur A., Madan A., Owen M.P., Ratcliffe A., Shaffer T., Hood L.;
RT "Sequencing of human chromosome 14q24.3 region."
RL Submitted (May-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 539-847 AND 1397-1822 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95319502; PubMed=7596406;
RA Sherrington R., Rogaev E.T., Liang Y., Rogaeva E.A., Levesque G.,
RA Ikeda M., Chi H., Lin C., Li G., Holman K., Tsuda T., Mar L.,
RA Foncin J.-F., Bruni A.C., Montesi M.P., Sobri S., Rainero I.,
RA Pinessi L., Nee L., Chumakov I., Pollen D., Brookes A.,
RA Sasseau P., Polinsky R.J., Masco W., da Silva H.A.R., Haines J.L.,
RA Petrick-Vance M.A., Tanzi R.E., Roses A.D., Fraser P.E.,
RA Rommens J.M., St George-Hyslop P.H.;
RT "Cloning of a gene bearing missense mutations in early-onset familial
RT Alzheimer's disease."
RL Nature 375:754-760(1995).
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1661.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AC007956; AAF61275.1; -.
DR EMBL: L40403; AAC42008.1; ALT_FRAME.
DR EMBL: L40400; AAC42006.1; -.
KW Nuclear protein.
FT DOMAIN 15 205 PRO-RICH.
FT DOMAIN 382 430 GLN-RICH.
FT DOMAIN 807 1209 ARG-RICH.
FT DOMAIN 1488 1577 ARG-RICH.
FT CONFLICT 621 621 P->S (IN REF. 2).
FT CONFLICT 1404 1404 T->I (IN REF. 2).
FT CONFLICT 1821 1821 K->E (IN REF. 2).
SQ SEQUENCE 1822 AA; 204947 MW; 8E6CB83FE240C7D2 CRC64;

```

Query Match 4.1%; Score 305; DB 1; Length 1822;
 Best Local Similarity 19.4%; Pred. No. 7.9e-05;
 Matches 333; Conservative 192; Mismatches 554; Indels 634; Gaps 84;

```

QY 139 PPLFLVPPRP-LSQRLQPRSRPILRQD---PVVFRYTVPTGCDTITAGS---S 191
Db 102 PPPPPPPPPPALSYQKQDQYKQMLNHRDGGPLVMELESPPESPVPVPPSGYMPSS 161
QY 192 DSGTSLRPASISISTNSQSLAPRVGNSPSAPAPVPDITATYVSIVHSEKSDGPPRD 251
Db 162 QSYMPPPPPPSYPPYPSISQYPL-PPAQPSQSGSPSSQSYLAPITPYSYSSSSSSQSYLSH 220

```


Db	863	EPGAAYVQKETEVS	SKGSAESPBGITTTTBE	GGCQDPEP	LEPVEKQGVNDIK	FDEG	922						
QY	1180	SKGSASSEK	---	RAKDREHSGSEKND	PKRKKA	---	QPKESTYDRLSEQGHFTLS 1232						
Db	923	AGFESESSAGYEKKAETEE	EAEEPEEDGEDV	SGSASHSPSTDE	ELAKADADVI	EKRR	982						
QY	1233	QS--SKERTSEKHE	SV--	RSSNKKDFT	PG--	RDKVDYDSRDXSSKRR	----- 1276						
Db	983	ESVANS	GDRAEDMD	EALEKEA	EQSEBE	EEEDKADAE	REEDPKTAE	DYVMAV 1042					
QY	1277	DERELARRK	-----	SPPR--	GKESISGCKSL	REERDLP	PKKGAESKNS	SSP--RDK 1327					
Db	1043	DKA	EAAGTDEQ	YVFLGPA	QPGVGS	RRP	PASSIH	ETLIP--GGSSE	EALASD	ENRED 1101			
QY	1328	KPHDKA	---	PLYTKR	PCETK	PVDK	NSGKER	KHA	EAHP	NGKSSG	CAN	CYLLTR 1380	
Db	1102	QPEFTAT	SGYGTQ	TIETIS	SEPT	PWDE	MS--	TPRDVMT	DET	NNEET	ESP	SQEFV	ITK 1157
RESULT	14												
ID	ATRX_MOUSE	STANDARD;	PRT;	2476	AA.								
AC	061687;												
DT	15-JUL-1999	(Rel. 38, Created)											
DT	15-JUL-1999	(Rel. 38, Last sequence update)											
DT	01-MAR-2002	(Rel. 41, Last annotation update)											
DE	Transcriptional regulator ATRX (X-linked nuclear protein)												
DE	(heterochromatin protein 2) (HPI alpha-interacting protein) (HPI-BP38 protein).												
GN	ATRX OR XNP OR HP1BP2.												
OS	Mus musculus (Mouse).												
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;												
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.												
OX	NCBI_Taxid	10090;											
RN	[1]												
RP	SEQUENCE FROM N.A.												
RX	MEDLINE	98213653;	PubMed	9545503;									
RA	Picketts D.J., Tastan A.O., Higgs D.R., Gibbons R.J.;												
RT	"Comparison of the human and murine ATRX gene identifies highly conserved, functionally important domains.";												
RL	Mamm. Genome	9:400-403(1998).											
RN	[2]												
RP	SEQUENCE OF 325-1176 FROM N.A.												
RX	MEDLINE	97133299;	PubMed	8978696;									
RA	le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H.,												
RT	Jeanmougin F., Lossos R., Chambon P.;												
RL	"A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors.";												
RN	[3]												
RP	SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC HETEROCHROMATIN.												
RX	MEDLINE	20040663;	PubMed	10570185;									
RA	McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,												
RT	Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,												
RL	Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;												
RT	"Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes.";												
RL	Proc. Natl. Acad. Sci. U.S.A.	96:13983-13988(1999).											
CC	-1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES												
CC	GENE EXPRESSION BY AFFECTING CHROMATIN.												
CC	-1- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEXIN V IN A CALCIUM AND												
CC	COSPHATIDYCHOLINE/PHOSPHATIDYLSERINE-DEPENDENT MANNER (By												
CC	similarity)												
CC	-1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC												
CC	HETEROCHROMATIN DURING INTERPHASE AND MITOSIS. PROBABLY BY												
CC	INTERACTING WITH HP1.												
CC	-1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.												
CC	-1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.												
CC	-----												
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration												
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -												

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
DR EMBL; AF026032; AAC08741.1; "-
DR EMBL; X99643; CAA67962.1; "-
DR MGD; MGI:103067; Xnp.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR InterPro: IPR001841; Znf_Ring.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HelicC; 1.
DR SMART; SM00184; Ring; 1.
DR DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;

KM	Zinc-finger.				
FT	ZN_FING	219	267	PHD-TYPE.	
FT	NP_BIND	1579	1586	ATP (POTENTIAL).	
FT	SITE	1704	1707	DEGH BOX.	
FT	DOMAIN	319	322	POLY-SER.	
FT	DOMAIN	735	738	POLY-SER.	
FT	DOMAIN	1001	1004	POLY-GLU.	
FT	DOMAIN	1130	1135	POLY-SER.	
FT	DOMAIN	1182	1185	POLY-SER.	
FT	DOMAIN	1238	1245	POLY-ASP.	
FT	DOMAIN	1484	1487	POLY-GLU.	
FT	DOMAIN	1924	1931	POLY-SER.	
FT	DOMAIN	2205	2208	POLY-LYS.	
FT	DOMAIN	2245	2248	POLY-GLU.	
FT	DOMAIN	2403	2408	POLY-GLN.	
SO	SEQUENCE	2476	AA: 278601	MM: 904A2B790FC4FF4C	CRC64:

Query match	4.18; Score 299; DB 1; Length 2476;
-------------	-------------------------------------

Best Local Similarity 19.0%; Pred. No. 0.00018;
Matches 292; Conservative 206; Mismatches 549; Indels 486; Gaps 68;

```

62 LICKDINTDAVVYIPCCGNSCDECI-----RTTLLBSDKHTCPTCHQNDVSPDALI 112
QY

```

Db 228 LICDF-----CHNAFCCKCILRLGRKELSTIMDENNQWYCYIC----QPEPLL 273

QY 113 -----ANKFLRQAVNNFKNETGYTKRLKQLPPLFLVPPBPRLSQRLQPRS 160

Db 274 DLTVCNSVFENEQLQOKKKIKVDSEKTSKYCDQTSKF---SPKKSSSSCGEEKK 329

QY 161 RSPILRQDDPVVERRYTVSPICSDTKT7AGSCSDSGTSLRLPAP-----SISLSLNSQ 212

```

Db      330 -----LESCS-----GSVSTYSHALSVPKEMIKTTLIETTSNMN 368

```

```

QY      213 S-----LAPVSGNPPSSAPAPVPDITATVSI-----SVHSEK 244
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db 369 SSYIKELQADNSEMTSAMKLCQLKSFKSVLDDIKKAHLAEEDLNSEIQALDDVHK 428

```

QY      245 SDGPFDSDNKLLPAAALTSSEKSGAS-----IAITAMEEKGVPG-----TS 288
      :  : |  :  : |  :  : |  :  : |  :  : |  :  : |  :  : |  :  : |

```

Db 429 NTKDKLSTD-----AKSEIRLKGKGSYSTEKREFLKLDARSSVKALDGEEOGRAHKSTS 482

```

289 PWNISFVGQSLHG-QLIPTGP-----VRINARPGG 320
      : | : | | | | | | | | | | | | | | |
290          : | : | | | | | | | | | | | | |

```

DB	483	GEHK----	GSGRKD6SQEPNTPEDLDMDIVSVPSVPEDIFDLSLSEAVEQSSADYQGD	539
001	3231	GRDGR	HWYRGLVIGDRCOTDDQDMDGCVDTGTDGDUUHGRRGDDMGCDSTDAIEDCV	3272

DY GRPWE---HSNKLGLVSPFQQLIRKGRSLCINSLNKGHHSHSRQRKYVSFSLPLALCFV 37

540 CNGCEDEET EESSGVIT NVSCSD-----SGDNISKVTAAYDKR-----EIV 580

00	378	PVPPPPV.PPBPBMT.PI.PGCVBPPDQFSGPQDPSQDDBTAGVSVBPBGPBPAPANT	STACFS	A37Z
DU	340	QNSGLEEESSGVKLVNVCQND	QNGNINQNV	IRAVAN
			ELFV	200

Db 581 KLTPTVSLNSNP-----IKGYDCQEFVSGEKGK-----S 609

[illegible]

between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: Z34289; CAB4063.1; -
DR EMBL: D21262; BAA04803.1; -
DR MIM: 602394; -
KW Nuclear protein; Phosphorylation; Repeat; GTP-binding; ATP-binding;
KW Alternative splicing.
FT DOMAIN 84 566 11 X 12 AA APPROXIMATE REPEATS OF AN
ACIDIC SERINE CLUSTER.
FT REPEAT 84 95 ACIDIC SERINE CLUSTER 1.
FT REPEAT 125 136 ACIDIC SERINE CLUSTER 2.
FT REPEAT 167 178 ACIDIC SERINE CLUSTER 3.
FT REPEAT 221 232 ACIDIC SERINE CLUSTER 4.
FT REPEAT 264 275 ACIDIC SERINE CLUSTER 5.
FT REPEAT 325 336 ACIDIC SERINE CLUSTER 6.
FT REPEAT 363 375 ACIDIC SERINE CLUSTER 7.
FT REPEAT 425 436 ACIDIC SERINE CLUSTER 8.
FT REPEAT 470 481 ACIDIC SERINE CLUSTER 9.
FT REPEAT 519 529 ACIDIC SERINE CLUSTER 10.
FT REPEAT 555 566 ACIDIC SERINE CLUSTER 11.
FT DOMAIN 68 82 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 204 382 INTERACTS WITH RPA194.
FT DOMAIN 384 587 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 601 617 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 563 563 PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT VARSPPLIC 241 241 K -> KWTITSVRAE (IN ISOFORM BETA).
FT CONFLICT 3 3 D -> A (IN REF. 2).
FT CONFLICT 133 133 R -> S (IN REF. 2).
FT CONFLICT 291 292 YA -> SV (IN REF. 2).
FT CONFLICT 456 456 S -> P (IN REF. 2).
SQ SEQUENCE .699 AA; 73720 MW; DFD4AD94EDF659FB CRC64;

Query Match 4.0%; Score 294; DB 1: Length 699;

Best Local Similarity 21.2%; Pred. No. 8.1e-05;
Matches 155; Conservative 121; Mismatches 279; Indels 176; Gaps 30;

QY 771 DATPDEPMDEASTTFKSVSDKREKDPVKSDTKRKSDGATAKKNV--LKPS 827
DB 3 DAGIRRVPSDLPLVGLRONQLESEVANKPAKATGATQO--DANASSLDIYFWLKSA 61
QY 828 KGPQEKVDGDRKSPSEPPPLKAKEEATKIDVAKPSSSQDEKVTGTPRAHKSARD 887
DB 62 KVPETRLQA-----NGPVAKKAKKASSDS--EDSSEEEVYQGPAP--KAAPV 108
QY 888 TTRQSQPTRRRKRKYPTKSS-----QKSPVPTRR--RPRSLRKINILIAEKNE 935
DB 109 ARVGLPPGKAATAKASESSSSSESRDDDEEDOKKOPVQGVKPPQ-----AKAARA 159
QY 936 REKRKKSVYKDFESSSMKISVETGTEIVKSPKRMGDEVKLETPPEKDIASSTP-- 993
DB 160 PPKAKKSSDSDSDSS-----EDPEPPKN-----QPKITPVYVKAQTAKAPPKP 202
QY 994 ---AKIKILNETGKKIGAENAATTKPESEKLESTSSK--IKQEKYKGA-----KR 1041
DB 203 ARAAPRIANGKAASSSSSSSSSSSSSDSEEEKAAATPKKTPVKQVAKAPVKAATTPTR 262
QY 1042 KYAGSGSGSSSTLVD-----YTSSTSGSGSPVRKS-----EKTDTKRTYIKTM 1084
DB 263 KSSSSSDSDSDSEDEOKKPMKKNPGDYSTAPPSPAPPKSLGTOPPKRAVEKQOPVESS 322
QY 1085 EEEYNDNTAPAD-----YIMI-----QVPQSKWMDKDFESEEDVKTQPI- 1127
DB 323 EDSDESDSSSEEEKKPTKAVYAKATTKPPAKKAASSSDSDSDSEDDAPSKPAG 382
QY 1128 ---QSYGKPSIITKNVTPSATAKY-----TEKESQPEKLOKL- 1164
DB 383 TTKNSNKPDAVVTTKSPAVKPAAPKOPVGGQKLTTRKADSSSSSEESSSEEEKTKKMY 442

QY 1165 ----PKASHHELM-----OHELRSKGSASSEKRAKDRHSGSEKNDPKRRKSGAO 1212
DB 443 ATTTPATKAAKLSLPKAPQPGSRDSDSDSSSEEEKTSKSAVKKPKQVAGGAA 502
QY 1213 PDKESTVDRLSPQGHFTLSQSKKETRTSEKHESVYG-----SSNKDFTPGDRDK 1262
DB 503 PSKPAS---AKKGAESSNSSSDSDSEEEKIKLKGSPRPQAPKANGTSALTAQNGX 558
QY 1263 VDYSDRDYSSSKRR-----DERGELARRK--DSPPRKESLSGQSKLREERDLPKGAES 1316
DB 559 AAKNSEEEEEEEKKAAVAVVSKSGSLKRRKQNEAAKAEETPOAKIKILOTPNTEPPKR--KKG 617
QY 1317 KKSNSPPR-----DKPHDKAPYETRRPC-----EETKPYDK--NSGKEREKHNA 1361
DB 618 EKRRASSPFRVREELIYDVRVADNS--FDARKGAAGDWGERANDLFTFKGSPRHEKT 675
QY 1362 EARNKGESSGA 1372
DB 676 KKKRGSYRGGS 686

Search completed: September 12, 2002, 19:49:53
Job time: 7233 sec

OM of: US-09-811-045a-1 to: EST:* out_format : pfs

Date: Sep 12, 2002 4:03 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODE=frame-p2n_model -DEV=xlh
-O/cgcn2.1/uspro.spool/US09811045/runat_12092002.132241.5977/app_query.fasta.1.1485
-DB=EST -OFMT=fastap -SUFFIX=est -GAOP=12.000 -GAPEXT=4.500
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAOP=4.500
-OGAPEXT=0.050 -XGAOP=10.000 -XGAPEXT=0.500 -FGAOP=6.000
-FAPEXT=7.000 -YGAOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXN=7.000 -STARR=1 -MATRIX=blonum62 -TRANS=humand0.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=2000000000 -USER=US09811045 -CGCN1.1.3800
-NCPU=6 -ICPU=3 -LONGLOG -DEV -TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLIPXY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-811-045a-1
Query length: 1404
Database: EST*
Database length: 13736207
Database length: -1841457050
Search time (sec): 2293.940000
```

score_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
gb_ntc:AF116625	+	2353.50	1812.40	1.1e-91	1742	AF116625 Homo sapiens PRO1155
gb_est1:AF1528817	-	1292.00	1002.18	1.4e-46	828	AF1528817 ms13308.x1 StrataGene
gb_est1:BI533873	+	1238.00	961.35	2.7e-44	759	BI533873 60317476F1 NCI CGAP
gb_est1:AU124466	+	1152.00	894.51	1.4e-40	832	AU124466 AU124466 NT2RM4 Homo
gb_est1:BF139148	+	1152.00	892.90	1.7e-40	1021	BF139148 60178418F1 NCI CGAP
gb_est1:AU124612	+	1118.00	866.44	4.0e-39	825	AU124612 AU124612 NT2RM4 Homo
gb_est1:BB626905	+	1111.00	865.00	6.2e-39	644	BB626905 BB626905 RIKEN full-16
gb_est1:BG575640	+	1100.00	854.64	2.3e-38	821	BG575640 60255870F1 NCI CGAP
gb_est1:BI686114	-	1080.00	838.65	1.8e-37	888	BI686114 60331310F1 NCI CGAP
gb_est1:BG575299	+	1073.00	833.89	3.3e-37	820	BG575299 60259795F1 NIH_MGC_87
gb_est1:BG522049	+	1050.50	816.34	3.2e-36	847	BG522049 60264665F1 NIH_MGC_75
gb_est1:BI333743	+	1044.00	812.08	5.5e-36	771	BI333743 60299921F1 NIH_MGC_12
gb_est1:BF720766	+	1042.00	810.82	6.5e-36	744	BF720766 m6b58602.y1 Soares.tny
gb_est1:BF117991	-	995.00	774.30	3.1e-35	678	BF117991 uz10e05.y1 NCI CGAP
gb_gss:AZ696947	-	985.00	766.87	1.8e-33	594	AZ696947 RPCI-23-238D2.TJ RPCI-
gb_est1:AF1322812	-	984.00	766.00	1.6e-33	592	AF1322812 ms13308.y1 StrataGene
gb_est1:BG087743	-	982.50	759.77	4.5e-33	577	BG087743 H3143308-5 NIA Mouse
gb_est1:BG168677	+	954.00	743.24	3.7e-32	737	BG168677 60231993F1 NIH_MGC_89
gb_est1:AU130966	+	951.00	740.53	5.3e-32	776	AU130966 AU130966 NT2RP3 Homo
gb_gss:AZ949647	+	936.00	730.62	1.9e-31	631	AZ949647 2M0213K12F Mouse 10kb
gb_est1:BE304165	+	917.00	716.61	1.1e-30	585	BE304165 60108585F1 NIH_MGC_12
gb_est1:BI334941	+	911.50	710.55	2.5e-30	739	BI334941 60299981F1 NIH_MGC_12
gb_est1:BE534832	+	911.00	712.34	2.0e-30	560	BE534832 601221396F1 NCI CGAP
gb_est1:BE281676	+	902.50	705.30	4.9e-30	597	BE281676 60109963F1 NCI CGAP
gb_est1:BM091010	+	896.00	699.90	9.7e-30	629	BM091010 1921b11.y1 Human Fetal
gb_est1:AF158608	-	885.00	692.01	2.7e-29	585	AF158608 w140604.x1 NCI CGAP
gb_est1:BI732070	+	876.00	663.48	8.0e-29	718	BI732070 60335285F1 NIH_MGC_94
gb_gss:AZ583889	+	865.00	675.05	4.4e-28	716	AZ583889 IM0388M03F Mouse 10kb
gb_gss:AZ090592	-	863.00	674.42	2.6e-28	638	AZ090592 RPCI-23-27B12.TJ RPCI-
gb_est1:AU131443	+	861.50	670.62	4.2e-28	893	AU131443 AU131443 NT2RP3 Homo
gb_est1:BF408911	-	857.00	670.23	4.4e-28	604	BF408911 UT-R-B52-bet-g-12-0-UT
gb_est1:AM209630	-	851.00	666.26	7.3e-28	555	AM209630 u145003.y1 Rasbass.mc
gb_est1:AF114571	+	835.50	652.85	4.1e-27	673	AF114571 HA1155 Human Fetal
gb_gss:AZ906558	+	827.50	647.05	8.6e-27	644	AZ906558 RPCI-24-164A17.TJ RPCI
gb_est1:AA664604	-	827.00	648.21	7.4e-27	529	AA664604 c045504.s1 NCI CGAP
gb_est1:BG661155	+	823.00	644.60	1.2e-26	566	BG661155 332994.MARC 1P16 Sus
gb_est1:AM988263	+	817.00	640.88	1.9e-26	505	AM988263 u904004.y1 Soares.mamm
gb_est1:BF408906	-	813.00	636.54	3.3e-26	594	BF408906 UT-R-B52-bet-g-07-0-UT
gb_est1:BF151531	+	811.00	636.29	3.4e-26	504	BF151531 uz116e03.y1 NCI CGAP

gb_est1:BE631744 + 807.00 632.88 5.3e-26 526 + BE631744 u055b11.y1 Soares.t
gb_est1:AL046980 - 803.50 627.12 1.1e-25 478 + AL046980 DKFZ586K0917.r1.58
gb_est1:AF1493321 - 792.00 622.08 2.1e-25 479 + AF1493321 t965c11.x1 Soares.N
gb_est1:BE567777 + 787.50 614.26 5.7e-25 835 + BE567777 601340443F1 NIH_MGC
gb_est1:BF921502 - 778.00 610.28 9.6e-25 547 + BF921502 MK2-NT0135-171100-0

seq_name: gb_ntc:AF116625

seq_documentation_block:

```
LOCUS AF116625 1742 bp mRNA linear HTC 08-MAY-2001
DEFINITION Homo sapiens PRO1155 mRNA, complete cds.
ACCESSION AF116625
VERSION AF116625.1 GI:7959751
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1742)
AUTHORS Zhang,C., Yu,Y., Zhang,S., Wei,H., Zhou,G., Ouyang,S., Luo,L.,
Bi,J., Liu,M. and He,F.
TITLE Functional prediction of the coding sequences of 121 new genes
deduced by analysis of cDNA clones from human fetal liver
2 (bases 1 to 1742)
UNPUBLISHED
AUTHORS Zhang,C., Yu,Y., Zhang,S., Wei,H., Zhou,G., Ouyang,S., Luo,L.,
Bi,J., Liu,M. and He,F.
TITLE Direct Submission
SUBMITTER Institute of Radiation Medicine, Department of Experimental Hematology,
Institute of Radiation Medicine, Beijing Taiping Road 27, Beijing
100850, P. R. China
```

FEATURES

location/Qualifiers

```
1..1742
/organism="Homo sapiens"
/db_xref="taxon:9606"
/codon_start=1
/codon_end="174438"
/tissue_type="liver"
/dev_stage="fetal"
713..961
/note="predicted protein of HQ1155"
/product="PRO1155"
/protein_id="AA71048.1"
/db_xref="GI:7959752"
/translation="MGEVLTLETLKKNATENGNGENSGMKNIKMLLEHSDPOQ
IERFLDRDCHLTSLGILPQEAETKLVKIVET"
```

```
BASE COUNT 647 a 385 c 342 g 368 t
ORIGIN
```

```
alignment_scores:
Quality: 2353.50 Length: 579
Ratio: 4.500 Gaps: 5
Percent Similarity: 90.328 Percent Identity: 77.720
```

alignment_block:
US-09-811-045a-1 x AF116625 ..

Align seg 1/1 to: AF116625 from: 1 to: 1742

```
367 SerProSerLeuPro...AlaThrProCysPheValProValProValPro 382
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
66 TCCTTTGATCCGCCCTCCCATACATCTCTCCCTCCGCGGCTTC 115
399 rGProProProProProProProHisThrLeuProLeuProProGlyValP 399
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
116 GTCCTCCACAGTTTCTCTCCAGTTTCTCTGCGGACGACACCCGCT 165
416 GlyTyrSerValProProProGlyPheProProAlaProAlaAsnIleSe 432
```

```

166 GGGATATGTCCTCCCTCAGGGTTCCTCAGCTCCTGCAATTTATC 215
432 rThAlaCysPheSerProGluValProThrAlaHisSerAsnThrMetp 449
216 AACCACTGGTATCATCAGAGTGCAGACAGCTCATTTCAATACCAATCC 265
449 rOThrPrGlnAlaProLeuLeuSerArgGluGluPheTyrArgGluGln 465
266 CAACAACAAGCAGCACCTTTGTCCAGGAGAAATCTATAGAGACAG 315
466 AsnAspLysGluArgGluSerLysPheProTyrSerGlySerTyrSe 482
316 CGACGACTAAAGAGAGTCTAAATCTCCATATGATGTTCTTCTGATATC 365
482 rArgSerSerTyrThrAspSerSerGlnGluLeuAlaGlnHisIleHis 498
366 AACAAATTCTATATCTTATCTAATCAAGATCTGTTCAACAGCTTAC 415
499 AlaLeuThrLeuSerProSerAlaAlaHisThrLeuAspLeuHisAs 515
416 GCTCTTATTCGATCATTCAGCGCTCAGCTCATCTCTCTATTCACGG 465
515 PHisProHisProGluGluGluAlaGluAlaArgSerAlaMetIleValH 532
466 TCACCTCCATACCCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 515
532 fAmetProAspLeuMetAspIleAlaHisIleAlaArgSerArgSerPro 548
516 ACGGTCTAGATCTCATGATATTCAT.CGATCTAGGTCAAGGTCCACCCCT 564
549 TyrArgArgTyrArgSerArgSerArgSerProGluPheArgGlyG 565
565 TACAGAGCTATCATCTCAGATCAAGATCTCCCAACGCTTAGGGACA 614
565 nSerProThrLysArgAsnValProArgGluGluLysGluArgGluTyrp 582
615 GTCTCCATAATTAACGTATATGCTCAAGGGGAAACAGAACGTAAATAT 664
582 heAsnArgTyrArgGluValProProProTyrAspIleLysAlaTyrTyr 598
665 TTATATATATCAGAGAAAGTTCACACCATATGATGATGAAGATATAT 714
599 GlyArgSerValAspPheArgAspProPheGluLysGluArgTyrArg 615
715 GGGAGAGTGTGACTTAGAGACCATTTGAAAGAAAGACCTACCGAGA 764
615 uTPrgLArgLysTyrArgGluTyrPyrGluLysTyrTyrLysGlyTyr 632
765 ATGGGAGAGAAATATAGAGAGTGTATGAAAAATATTATTAAGGTTATG 814
632 laValGlyAlaGlnProArgProSerAlaAsnArgGluAspPheSerPro 648
815 CTGCTGAGAGCAGACCTAGACCTCAGCAAAATAGAGAAATTTTCTCCA 864
649 GluArgLeuLeuProLeuAsnIleArgAsnSerProPheThrArgGly 665
865 GAGAGATTTTGGCCACTTAACATCAGAAATCTCCCTTCAACAAAGGCGC 914
665 rArgGluAspPyrAlaIleGlnGlnSerHisArgAsnArgGlnGlyG 682
915 CAGGAGAGACTATGTTGGTGGCAAGTCAATAGAGTCAAGGAAACATAG 964
682 LysAsnTyrProGluLysLeuSerThrArgAspSerHisAsnAlaLysAsp 698
965 GCAACTTCCAGAAAGCTTTCAACAGAGATGTCTCAATCAACAAGAGAT 1014
699 AsnProLysSerLysGluLysGluSerGluAsnValProGlyAspGly 715
1015 AATCAACAAGTCAAAAGAGAGAGAGTGAAGAAACCTCCAGAGATGTGTA 1064
715 sGlyAsnLysHisLysLysHisArgLysArgArgAsnGluLysGlyG 732

```

```

1065 AGGAATATAGCATATAGAACACAGAAAAAGAGA.....AAAGGG 1105
732 lUglUserGluSerPheLeuAsnProGluLeuLeuGluThrSerArgLys 748
1106 AGGAATGTGAGGTTTCTGACCCAGAGTTATTAGAGACTTCAGGAAA 1155
749 CysArgGlySerSerGlyIleAspGluThrLysThrAspThrLeuPhe 765
1156 TCAGAGAACCTACAGCTGTGAAGAAATTAACACATCATATGTTGT 1205
765 lLeuProSerArgAspAspAlaThrProValArgAspGluProMetAsp 782
1206 TCTCCCAAGTATAGATATGTCACACCTGTTAGAGATGAACCAATGATG 1255
782 lAglUserIleThrPheLysSerValSerAspLysAspLysArgLys 798
1256 CAGAAATCATCTTTTAATCACTGTCTGAAAAAGACAGAGAGAAAG 1305
799 AspLysProLysValLysSerAspLysThrLysArgLysSerAspLys 815
1306 GATTAACCAAAAGCAAGGAGTATTAACCAAGCAAGCAATGATGATC 1355
815 rAlaThrAlaLysLysAspAsnValLysLysProSerLysGlyProGln 832
1356 TCGTGTCTCCAAAAAAGAAATATTGTAACCTGCTTAAGGAGCCCAAG 1405
832 lUlysValAspGlyAspArgGluLysSerProArgSerGluProLeu 848
1406 AAAAAGTATGAGAGAAAGTGAAGATCTCTCGATCTGAACCTCAAT 1455
849 LysLysAlaLysGluGluAlaThrLysIleAspSerValLysProSer 865
1456 AAAAAGCCAAAGAGAGACTCCGAGACGCAATCTAATATCATATC 1505
865 rSerSerGluLysAspGluLysValThrGlyThrProArgLysAlaHis 882
1506 TTCTCTCAGAGATTAATAATCACTGCAACCCCAAGAAACCTCACT 1555
882 eLysSerAlaLysAsp.ThrArgArgGlnSerGlnProArgThrArg 898
1556 CTAAATCAGCAAAAGAACCAACCAAAACCAAGCAAAAGAGAGAAA 1605
898 gSerLysArgThrValProLysThrSerSerGlnLysSerGlnProVal 915
1606 GTGAAGAGAGCACTATTCAAAGATGCAAAACAGAAAGCTAACACT 1655
915 rGThrArgArgProArgSerLeu..ArgLysIleAsnTyrLeuIleAla 930
1656 GGAAGAAAGGCCAAGAGCTTAATGAGAAAAACAAACCACTGATATA 1705
931 ArgGluLysAsnGluArgGluLysArgLysLys 941
1706 AGGAGAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1738

seq_name: gb_est1:A1528817
seq_documentation_block:
LOCUS      A1528817              828 bp      mRNA      linear      EST 16-APR-1999
DEFINITION ms13908.x1 Stratiogene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:606878 3' similar to TR:P97868 P97868 RETINOBLASTOMA BINDING
PROTEIN 6 ;, mRNA sequence.
ACCESSION  A1528817
VERSION    A1528817.1 GI:4442952
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus.
REFERENCE  1 (bases 1 to 828)
AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
B., Swaller,T., Gibson,M., Pape,D., Harvey,N., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,

```



```

/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH108"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NH"

```

```

BASE COUNT      253 a      178 c      213 g      115 t
ORIGIN

```

```

alignment_scores:
  Quality: 1238.00      Length: 249
  Ratio: 5.012          Gaps: 2
  Percent Similarity: 99.197      Percent Identity: 98.394

```

```

alignment_block:
US-09-811-045A-1 x BI453873 ..

```

```

Align seg 1/1 to: BI453873 from: 1 to: 759

```

```

1133 ProSerSerIleIleLysAsnValThrLysProSerAlaThrAlaLys 1149
      |||||||
3  CCATCGAGTATATATAAAATGTCACTACTAAGCCATCGCTACGCGCTAA 52

1149 sTyrThrGluLysGluSerGluGluProGluLysLeuGluNlyLeuProL 1166
      |||||||
53  GTACACCGAAGAGAAAGGAGAGAGAGCCGACGAGAAACGCGAAGCTTCCCA 102

1166 ySGLuAlaSerHisGluLeuMetGluHisGluLeuArgSerSerLysGly 1182
      |||||||
103  AGGAGAGCGAGCCAGCGATGATGACAGACAGACTCTCAGAGCTCAAAAGGCG 152

1183 SerAlaSerSerGluLysGlyAlaLysAspArgGluHisSerGlyLys 1199
      |||||||
153  AGTGGCTCCAGTACAGAGAGGCGAGAGCAAGGAGCCGAGGAGCTCAGGCTC 202

1199 rGluLysAspAsnProArgLysArgLysSerGlyAlaGluProAspLysG 1216
      |||||||
203  GGAGAGAGGCAACCTGACAAAGAGAGAGCGGTCGCCGACAGACAAAG 252

1216 LuSerThrValAspArgLeuSerGluGluGlyHisPheLysThrLeuSer 1232
      |||||||
253  AGAGACACGTGGAGCCGCTGATGAGAGGAGCATTTTAAGACTCTCTCT 302

1233 GluSerSerLysGluThrArgThrSerGluLysHisGluSerValArgL 1249
      |||||||
303  CAGTCTTCCAAAGAGACAGGAGCTCAGAGAGCAGAGTCTGTCTGTGG 352

1249 YSerSerAsnLysAspPheThrProGlyArgAspLysLysValAspLys 1266
      |||||||
353  TTCTCAATATAAGACTCTCTCTCCGCGGCAAAAGAGTCTGTCTGGGCA 402

1266 sPserArgAspTyrSerSerSerLysArgArgAspGluArgLysGluLeu 1282
      |||||||
403  ACAGAGAGGATTAATTCAGTTCCAGCCAGAGAGAGAGAGAGTGAATTA 452

1283 AlaArgArgLysAspSerProProArgLysGluSerLeuSerGlyLys 1299
      |||||||
453  GCAGAGAGAGAAAGACTCTCTCTCCGCGGCAAAAGAGTCTGTCTGGGCA 502

1299 nLysSerLysLeuArgGluArgAspLeuProLysLysGlyAlaGluLys 1316
      |||||||
503  GAAAAGCAAGCTGAGGAGAGAGAGATTACCTAAAAAGGGGCGCGAGT 552

1316 eLysLys..SerAsnSerSerProProArgAspLysLysProHisAspH 1332
      |||||||
553  CAAACAAACAAGTAATTTCTAGCCCCCAAGAGACAAAAAGCCTCATGATC 602

1332 IsLysAlaProTyrGluThrLysArgProGlyGluGluThrLysProVal 1348
      |||||||

```

```

603 ATMAAGCCCCCTACGAAACTAAACGCCCATGTGAAGAGACAAAGCCTGTA 652
1349 AspLysAsnSerGlyLysGluArgGluLysHisAlaAlaGluAlaArgAs 1365
      |||||||
653  GATTAATAACTCTGGGAGAGAGCGGAGAGCATGTGCTGTAAGCTCGGCA 702

1365 nGlyLysGluSerSerGlyAlaAsnGlyHisValTyrLeuThr 1379
      |||||||
703  TCGGAAGAGTCCAGTGTGGTGGCAAAATGCCATGTAATTAACCTAAC 746

seq_name: gb_est1:AUI24466

```

```

seq_documentation_block:
LOCUS      AUI24466      832 bp      mRNA      linear      EST 23-OCT-2000
DEFINITION AUI24466 NT2RM4 Homo sapiens cDNA clone NT2RM400050 5', mRNA
sequence.
ACCESSION  AUI24466
VERSION    AUI24466.1 GI:10949182
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 832)
            Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Salto,K., Yamamoto,J.,
            Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
            Isogai,T.
            HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
            Salto,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
            ,Y., Sugano,S., Isogai,T.)
            Unpublished (2000)
COMMENT    Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.

```

```

FEATURES
    source
        1..832
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="NT2RM400050"
            /clone_lib="NT2RM4"
            /cell_type="teratocarcinoma"
            /cell_line="NT2"
            /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
            precursor cells"

```

```

BASE COUNT      261 a      199 c      144 g      225 t      3 others
ORIGIN

```

```

alignment_scores:
  Quality: 1152.00      Length: 278
  Ratio: 4.760          Gaps: 3
  Percent Similarity: 87.050      Percent Identity: 83.453

```

```

alignment_block:
US-09-811-045A-1 x AUI24466 ..

```

```

Align seg 1/1 to: AUI24466 from: 1 to: 832

```

```

1 MetMetGluValLysAspProAsnMetLysGlyAlaMetLeuThrAsnTh 17
      |||||||
53  ATGATGGAAGTGAAGATCTTAATATGAAGGTGCAATGCTTACCAACAC 102

17 rGlyLysTyrAlaIleProThrIleAspAlaGluAlaTyrAlaIleGlyL 34
      |||||||
103  TGGAAATATGCAATACCACTATGATGACAGAGCATATATGCAATTGGAA 152

```

seq_name:	gb_est2:BF139148
seq_documentation:	1021 bp mRNA linear EST 24-OCT-2000
LOCUS	BF139148
DEFINITION	60178418BF1 NCI_CGAP_Lu30 Mus musculus cDNA IMAGE:4012269 5'
ACCESSION	mRNA sequence.
VERSION	BF139148
KEYWORDS	BF139148..1 GI:10978188
SOURCE	EST.
ORGANISM	house mouse. Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 1021)	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus			
	NIH-MGC	http://mgc.nci.nih.gov/		
	National Institutes of Health, Mammalian Gene Collection (MGC)			
	Unpublished (1999)			
	Contact: Robert Strausberg, Ph.D.			
	Email: cgapds-remail.nih.gov			
	ISSUE Procurement: Gilbert Smith, Ph.D.			
	cDNA Library Preparation: Life Technologies, Inc.			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
	DNA Sequencing by: Incyte Genomics, Inc.			
	clone distribution: NCI-CGAP clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LLNL at:			
	http://image.llnl.gov			
	Plate: L14W9252 row: O column: 22			
	High quality sequence loc: 665.			
	location/Qualifiers			
	1. 1021			
	/organism="Mus musculus"			
	/strain="C57BL/6J"			
	/db_xref="taxon:10090"			
	/clone_image="IMAGE:4012269"			
	/clone_id="NCI-CGAP-Lu30"			
	/tissue_type="tumor, metastatic to mammary"			
	/lab_host="DH10B"			
	/note="Organ: lung; Vector: pCMV-SPORT6; Site: 1; NotI;			
	Site 2: SalI; transgenic model WNT-1, expression driven by			
	MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo			
	dr. Library constructed by Life Technologies.			
	Investigator providing samples: Gilbert Smith, NIH			

```

alignment_scores:
  Quality: 1152.00
  Ratio: 3.752
  Percent Similarity: 78.517
  Percent Identity: 70.077

alignment_block:
  US-09-811-045A-1 x BF139148 ..

Align seg 1/1 to: BF139148 from: 1 to: 1021

937 GtLysArgLys...LysSerValAspLysAspPheGluSerSerSerme 952
|||||
1 GAAAAACGGAGACAAAGAGTGTAGATAAAGATTGTGAGCTGCTCAAT 50

952 tLysIleSerLysValGluGlyThrGluIleValLysProSerProLysA 969
|||||
51 GAAAACTCTAAAGTGAAGGAACAAAAATAGTGAACCATCCCAAAAC 100
|||||
969 rGlyMetGluGlyAspValGluLysLeuGluThrProGluLysAsp 985
|||||
101 GGAATAATGGAGGTGATGTGAAGAACTGGAAGGCCCAAGAAAGAGAC 150
|||||
986 LysIleAlaSerSerThrThrProAlaLysLysIleLysLeuAsnArgI 1002
|||||
151 AAGATTGCATCATCACTACTCTCAGCCAAAAAATCAAACTCAACAGAGA 200
|||||
1002 uThrGlyLysLysIleGlyAsnAlaGluAsnAlaSerThrThrLysGluP 1019
|||||
201 AACTGCAAAAAAAATGTGAAATGCACAAAAATGTCATCTACTACAAAAAGAC 250
|||||
1019 rSerSerLysLeuGluLysSerThrSerSerLysIleLysGlnGlyLysVal 1035
|||||
251 CCTCTGAAAAAATTGGAGTCACACTTCAGCAAAATCAAAACGAGAAAAAGTC 300
|||||
1036 LysGlyLysAlaLysArgLysValAlaGlySerGluGlySerSerSerTh 1052
|||||
301 AAGGGAAGGCCCAACGAAAGTACGTGGGTCCAGAAAGCCCTCCAGCTCCAC 350
|||||
1052 rLeuValAspIlyrThrSerThrSerSerThrGlyLysSerProValArgL 1069

```

```

|||||
351 GCTTGGGATTACACGATGACAGTTCACTGAGGACGACTCTGTAGCA 400
1069 ysserGlululysThrAspThrLysArgThrValIleLysThrMetGlu 1085
|||||
401 AATCTGAAGAAAAGACGATACAAAGCAAGTCATTAAACTATGAGG 450
1086 GluTyrAsnAsnAspAsnThrAlaProAlaGluAspValIleIleMetI 1102
451 GAATATTAATGAATATACACAGCTCTGCTGAAGATGTTATATATGAT 500
1102 eGlnValProGlnSerLysTrpAspLysAspAspPheGluSerGluGlu 1119
501 CCAGGTTCTCTCAGTCCCAATGGATMAAGATGACTTTGAGTCTGAAGA 550
1119 LuAspAluLysThrThrGlnProIleGlnSerValGlyLysProSerSer 1135
551 AAGATGTTAAAAAGCACACAGCTGTACAGAGTGTAGGAACCATCGAGT 600
1136 IleIleLysAsnValThrThrLysProSerAlaThrAlaLysTyrThr 1151
601 ATTATATAAAATGTCACACTACTTAAGCATCGGGTTACGGCTAAGTACACT 650
1152 GluLysGluSerGluGlnProGluLysLeuGlnLysLeuProLysGlu 1167
651 GAGAAAGAAAGCAGACAGACAGCACTGCAGAACTCCCAAGGAG 700
1168 AlaSerHisGluLeuMetGlnHisGluLeuArgSerSerLysGlySerAl 1184
701 GCGAGCAGAGTGTGATGCGACGACGATCAGAGCTCAAGGCGCCG 748
1184 aserSerGluLysGlyArgAlaLysAspArgLuhisSerGlySerGlu 1200
749 GTCCAGTGTGAAGGCGGAGGCA...GAGCGGAGACACTCAGTCCGACA 795
1201 LysAspAsnProAspLysArgLysSerGlyAlaGlnProAspLysGlu 1217
796 AGG...CAACCTGACAGAGAAAGGCGCCAGCAAGAAAGGAGCGCG 842
1217 rThrValAspArgLeuSerGluGlnLysPheLysThrLeuSerGlns 1234
843 GACG...|||||
1234 eSerLysGluThrArgThrSerGluLysHisGluSerValArgLysSer 1250
848 CGAGTGAAGCGGAGATACAGAACCCCTCAGCTCCAGAACAGGATCG 897
1251 SerAsnLysAspPheThrProGluArgAspLysLysValAspLysPse 1267
898 AAAAACCAAG...|||||
1267 rArgAspTyrSerSerSerLysArgArgAspGluArgGlyLueLueAla 1284
907 .....TGGCAGGCGCCCAAAAAAGACCG..... 930
1284 rGArgLysAspSerProArgGlyLysGluSerLeuSerGlyLys 1300
931 .....GGGAAAAACAAGAGGCGCCACAGAGA 957
1301 SerLysLeuArgGluGluArgAspLeuProLysLysGlyAlaGluSerL 1317
958 TCACGCGCGAGAGAAACAAGAAATCGCAAGAAATCCCGAGACAAATACGA 1007
1317 sLysSerAsnSer 1321
1008 CAAGCGCAACACC 1020
seq_name: gb_est1:AU124612
seq_documentation_block:
LOCUS AU124612 825 bp mRNA linear EST 23-OCT-2000
DEFINITION AU124612 NT2RM4 Homo sapiens cDNA clone NT2RM4000274 5', mRNA
sequence.

```

```

ACCESSION AU124612
VERSION AU124612.1 GI:10949328
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 825)
AUTHORS Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
Salto,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
,Y., Sugano,S., Isogai,T.)
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
Source Location/Qualifiers
1..825
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="NT2RM4000274"
/cell_type="teratocarcinoma"
/note="Vector: pME18SFT3; mRNA from uninduced NT2 neuronal
precursor cells"
BASE COUNT 258 a 198 c 143 g 223 t 3 others
ORIGIN
alignment_scores:
Quality: 1118.00 Length: 276
Ratio: 4.658 Gaps: 1
Percent Similarity: 86.957 Percent Identity: 83.333
alignment_block:
US-09-811-045A-1 x AU124612 ..
Align seg 1/1 to: AU124612 from: 1 to: 825
1 MetMetGluValLysAspProAsnMetLysGlyAlaMetLeuThrAsnTh 17
|||||
53 ATGATGAGAAAGTGAAGATCTTAATATGAAAGGTGCAATGCTTACCAACAC 102
17 rGlyLysTyrAlaIleProThrIleAspAlaGluAlaTyrAlaIleGlyL 34
|||||
103 TCGAAATATATGCAATACCACTATAGATGCACAGACATATGCAATTTGGA 152
34 ySLysGluLysProPheLeuProGluGluProSerSerSerSerGlu 50
|||||
153 AGAAGAGAAACCTCCCTTCTTACAGAGAGCCATCTTCTTCCACAGAA 202
51 GluAspAspProIleProAlaGluLeuLeuLysLeuIleCysLysAspDl 67
|||||
203 GAAAGATGATCTTATCCAGATGAATTTGTTGTCTCATCTGCAGAGATAT 252
67 eMetThrAspAlaValIleProCysCysGlyLysAsnSerSerCysAspG 84
|||||
253 TATGACTGATGCTGTGTGATTCCTGCTGCTGAGAAACAGTACTGTGATG 302
84 LuCysIleArgThrThrLeuLeuGluSerAspLysHisThrCysProThr 100
|||||
303 AATGTATTAAGAACAGCACTCTCGAATCAGATGACACACATGTCTCGAGC 352

```

101 CysH1sg1nAsnAspValSerProspAlaLeuIleAlaAsnLysPhe 117
117 uArg1nAlaValAsnAsnPhelLysAsnG1uThrGlyTrpThrLysArgL 134
403 ACGAAGAGCGCTGTAATTAACCTCAAAATAATGAACCTGGCTATACAAAAAGAC 452
134 euArg1LysG1nLeuProProPheLeuPheLeuValProProProArgPro 150
453 TACGAAACAGTTACCTCCTCCACACACCCCAATACACACCTCCGACACCA 502
151 LeuSerG1nArgAsnLeuG1nProArgSerArgSerProL1leuArgL 167
503 CTGATTCCAGAGAACCTACACACCTCTGATGAATCTCCGATATCAAGACA 552
167 nG1nAspProValValPheArgTrpThrValSerProThrCysSerAspT 184
553 ACAAGATCCTCTTATGATTCACAGTACATCATTCATCACT 592
164 hLyThrAlaLeuLysCysSerAspSerG1TrpThrLeuSerArgLeuPro 200
593CACCA 598
201 AlaProSerL1leSerSerLeuThrSerAsnG1nSerSerL1leuAlaProP 217
599 GCTCGCTCATATCTTCATTAACCTCTATCAAGCTCTTCCTGGCCCTCC 648
217 oValSerG1nAsnProSerSerAlaProAlaProValProAspL1leThra 234
649 TGTCGTCTGAAATCCGCT.TCTGCTCCACCTCCTGTT.CCTGATATAAGTG 696
234 laThValSerL1leSerValHisSerG1uLysSerAspGlyProPheArg 250
697 CAACAGTATCCATTCAGTTCATTCAGAAATAATCAATGAGCATTTCGNG 746
251 AspSerAspAsnLysLeuLeuProAlaAlaAlaLeuThrSerG1uHis 267
747 GATTCTGATATTAATAATATTGCCAGCTGCAGCTCTTGCAATCAGACGACT 796
267 rLysG1yAlaSerSerL1leAlaL1leThr 276
797 AAAGGNAACCTCTTNAATGCAATTACC 824
seq_name: gb_est1:BB626905
seq_documentation_block:
LOCUS BB626905 644 bp mRNA linear EST 26-OCT-2001
DEFINITION BB626905 RIKEN full-length enriched, 12 days embryo, embryonic body
between diaphragm region and neck Mus musculus cDNA clone
943004J02.5', mRNA sequence.
ACCESSION BB626905
VERSION BB626905.1 GI:16464746
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 644)
AUTHORS Arakawa,T., Carlinici,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
'M., Himoto,K., Horii,F., Iehi,Y., Ito,M., Kawai,J., Koudou,H., Koudou,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasakih,H.,
'D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
'Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
COMMENT Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@gsic.riken.go.jp,
URL: http://genome.gsic.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujikake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura
S., Kawaji, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.

Riken integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
Y., and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001).

Kondo, S., Shinagawa, A., Salto, T., Kiyoasa, H., Yamanaoka, I., Alzawa
K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsic.riken.go.jp>) for
further details.

e mouse tissues.

FEATURES

SOURCE

Location/Qualifiers

I . 644

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="9430040J02"

/clone_1bp="Riken full-length enriched, 12 days embryo,

embryonic body between diaphragm region and neck"

/tissue_type="embryonic body between diaphragm region and

neck"

/dev_stage="12 days embryo"

/lab_host="DH10B"

/note="Site_1: Salt; Site_2: BamHI; CDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in

RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was

primed with a primer [5'

GAGACGAGAAGATCCACAAGCTCTTTTGTGGTTTTTTVN 3'] , cDNA was

prepared by using trehalose thermo-activated reverse

transcriptase and subsequently enriched for full-length by

cap-trapper. cDNA went through one round of normalization

to Rot = 20.0 and subtraction to Rot = 370.0. Second

strand cDNA was prepared with the primer adapter of

sequence [5' GAGACGAGAAGTCCTCGATTAAATAAATTAATCCCCCCCCCC

3']. cDNA was cleaved with XhoI and BamHI."

BASE COUNT 215 a 150 c 188 g 91 t

ORIGIN

alignment_scores:

quality:	1111.00	length:	214
ratio:	5.192	gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-09-811-045A-1 x BB626905 ..

Align seg 1/1 to: BB626905 from: 1 to: 644

```

1143 LysPProSeAlatThrAlaLysTrpThrGlutyluSGluSerClnglnPrcoI 1159
|||||.....
2  AAGCCATCGGCTTAGTGACACCAGGAAGGAAGGACAGCACGCCCA 51
```

```

1159 uLysLeuGlnLysLeuProLysGluAlaSerHisGluLeuMetGlnHis 1176
|||||
52 GAAGACTGCAAGCTTCCCAAGGAGGAGCCAGCAAGCTGATGAGCAGC 101
|||||
1176 LuLeuArgSerSerLysGlySerAlaSerSerGluLysGlyAlaGlyAlaLys 1192
|||||
102 AGCTCAGGAGCTCTCAAGGCGAGTGCCTCCAGTGAAGAGGCGCAGGCCAAG 151
|||||
1193 AspArgGluHisSerSerGlySerGluLysAspAsnProAspLysArgLysSe 1209
|||||
152 GACCGGAGAGCACTCAGAGTGGAGAGAGACACCTTACACAAGAGAGAG 201
|||||
1209 rGlyAlaGlnProAspLysGluSerThrValAspArgLeuSerGluGlnG 1226
|||||
202 CGGTGCCAGCCAGACAAAGAGAGCACTGTGACCCGCTGATGAGCAGC 251
|||||
1226 LysHisPheLysThrLeuSerGlnSerSerLysGluThrArgThrSerGlu 1242
|||||
252 GACATTTTAAAGACTCTCTCAGTCTTCCAAAGAGACCAAGACTTCCAGAG 301
|||||
1243 LysHisGluSerValArgGlySerSerAsnLysAspPheThrProGlyArg 1259
|||||
302 AAGCAGCAGAGCTGTCTGCTGCTCTCAATAAAGACTTCACTCTGCTAG 351
|||||
1259 gaAspLysLysValAspTyrAspSerArgAspTyrSerSerSerLysArg 1276
|||||
352 AGACAAAGAAAGTGAGCTACGACAGCAGGATTATTCAGTTCCAAAGCAA 401
|||||
1276 rgaAspGluArgGlyGluLeuAlaArgArgLysAspSerProProArgGly 1292
|||||
402 GAGACGAGAGAGTGAAATTAGCAAGAGAAAGACTCTCTCCCGGGGGC 451
|||||
1293 LysGluSerLeuSerSerGlyGlnLysSerLysLeuArgGluGluArgAspLe 1309
|||||
452 AAAGAGCTCTGTCTGGCGAGAAAGCAAGTGGAGGAGAGAGAGATT 501
|||||
1309 uProLysLysGlyAlaGluSerLysLysSerAsnSerSerProProArg 1326
|||||
502 ACCTAAAGAGGGGGCGAGTCAGTCAAAAAAGTAATTCAGCCCGCCAGAG 551
|||||
1326 spLysLysProHisAspHisLysAlaProArgGluThrLysArgProCys 1342
|||||
552 ACAAAAAGCCCATGATCATTAAGCCCCCTAGCAAGCTTAAGCGCCATGT 601
|||||
1343 GluGluThrLysProValAspLysAsnSerSerGlyLysGluArg 1356
|||||
602 GAAGAGCAAGAGCTGTAGTAAAGTCTGCGAAGAGCGG 643
|||||
seq_name: gb_est2:BG575640

seq_documentation_block:
LOCUS BG575640 821 bp mRNA linear EST 10-APR-2001
DEFINITION 602598704F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4707481 5',
mRNA sequence.
ACCESSION BG575640
VERSION BG575640.1 GI:13583293
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 821)
NIH-MGC http://mhc.nci.nih.gov/.
Mammalian Gene Collection (MGC)
National Institutes of Health, Mammalian Gene Collection (MGC)
unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: rpe i.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

```

```

found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10577 row: g column: 02
High quality sequence stop: 802.
FEATURES
    source
        1..821
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4707481"
            /clone_lib="NIH_MGC_87"
            /tissue_type="mammary adenocarcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: breast; Vector: pCMV-Sport6; Site: 1: Not;
            Site: 2: Salt; Cloned unidirectionally; oligo-dT primed.
            Average insert size 1.383 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH-MGC Library."
BASE COUNT      206 a      252 c      156 g      205 t      2 others
ORIGIN
alignment_scores:
    Quality: 1100.00      Length: 285
    Ratio: 4.583      Gaps: 3
    Percent Similarity: 84.211      Percent Identity: 79.649
alignment_block:
    US-09-811-045A-1 x BG575640 ..
Align seg 1/1 to: BG575640 from: 1 to: 821
134 LeuArgGlyGlnLeuProProPheLeuPheLeuValProProArgPro 150
|||||
24 GTCCGAAACAGTTACTCTCTCCACACCCCAATACCTCCGAGACC 73
|||||
150 oLeuSerGlnArgAsnLeuGlnProArgSerArgSerProIleLeuArg 167
|||||
74 ACTGATTCAGAGAGACCTACACCTCGATGAGATCTCCGATATCAGAG 123
|||||
167 InGlnAspProValAlaPheArgTyrThrValSerProThrCysSerAsp 183
|||||
124 AACAAGATCCTCTTATGATTCAGTACATCTTCATCACT..... 164
|||||
184 ThrLysThrAlaGlySerCysSerAspSerGlyThrLeuSerArgLeuPro 200
|||||
165 .....CACCC 169
|||||
200 oAlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAlaPro 217
|||||
170 AGTCCTGCTATATCTTCACTTAATCACTTCTTCCCTGGCCCTC 219
|||||
217 roValSerGlyAsnProSerSerAlaProAlaProValProAspIleThr 233
|||||
220 CTGTGCTCGAATATCCGCTCTGCTCAGCTCCTGATGATTAAT 269
|||||
234 AlaThrValSerIleSerValHisSerGluLysSerAspGlyProPheArg 250
|||||
270 GCAACAGTATCCATATCAGTTCAATTCAGAAATCAAGATGAGACCTTTG 319
|||||
250 gaSPSerAspAsnLysLeuLeuProAlaAlaAlaLeuThrSerGluHis 267
|||||
320 GGATTCGATATATAAATATGCGCAGCTGCGACTTGACATCAGAGCACT 369
|||||
267 erLysGlyAlaSerSerIleAlaIleThrAlaLeuMetGluGluGly 283
|||||
370 CAAGGAGACCTCCTCAATTCATACCGCTTATGAGAGAGAGAGGT 419
|||||
284 ValProGlyThrSerProThrPasnSerIlePheValGlyLysLeuLeu 300
|||||
420 TAACAGGTGCTGTTCTTGGAAACCCCATCTTGTGAGACAGTCATTAT 468
|||||
300 euHisGlyGlnLeuIleProThrGlyProValArgIleAsnAlaAla 316
|||||

```

469 TGCATGACAGTTGATCCCAACAGTGTCCAGTAAAGATTAATTAATCTGCT 518
 317 ATGProGlyGlyArgProGlyTTPGluHisSerAsnLysLeuGlyTy 333
 519 CGTCCAGGTGGTGTGACACAGGTGGAGACATTCACAACTGCTGCTA 568
 333 LLeuValSerProGlnGlnLeuArgArgGlyLysArgSerCysTyra 350
 569 TGTGGTTTCCAGCACAAATTAAGAGAGGAGAGAGAGAGAGAGAGAG 618
 350 rSerLeuAsnArgGlyArgHisHisSerGluArgSerGlnArgThrGln 366
 619 GAGGTAAACCGTGGCGGACACAGCAGCAGCAGCAGCAGCAGCAGCAG 668
 367 SerProSerLeuProAlaThrProCysPheValProValProProPro 383
 669 GGGCCCGCATACACAGCACTGCTTGTACTGTCTCCACACT. CC 717
 383 oLeuTyProProProProHisThrLeuProLeuProProGlyValPro 400
 718 TTTGTATCCCGCTCTCCATACACTTCTCTCCCGGCGGTCTCTC 767
 400 roProGlnPheSerProGlnPheProSerSerGlnProProThrAlaG 416
 768 CTCACACAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 817
 416 y 416
 818 T 818

seq_name: gb_est2:BI686114

seq_documentation_block:

LOCUS BI686114 888 bp mRNA linear EST 18-SEP-2001
 DEFINITION 603313103F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5352960 5',
 mRNA sequence.

ACCESSION BI686114

VERSION BI686114.1 GI:15648742

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 888)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgep@strausberg.nih.gov

Tissue Procurement: Jeffrey Green M.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1A11897 row: f column: 01

High quality sequence stop: 757.

Location/Qualifiers

1. 888

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:5352960"

/clone_1ib="NCI_CGAP_Mam6"

/sex="female, virgin"

/tissue_type="infiltrating ductal carcinoma"

/dev_stage="5 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigator

providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 211 a 183 c 255 g 239 t
 ORIGIN

alignment_scores:
 Quality: 1080.00 Length: 311
 Ratio: 4.320 Gaps: 5
 Percent Similarity: 80.386 Percent Identity: 74.277

alignment_block:

US-09-811-045A-1 x BI686114/rev ..

Align seg 1/1 to reverse of: BI686114 from: 1 to: 888

74 TLeProCysCysGlyAsnSerSerCysAspGluCysIleArgThrThrLe 90
 880 ATTCCCTGCTGTGGAGACAGTACGTGTGATGAAGGTAGAGAGACAGAT 831
 90 uLeuGluSerAspLysHisThrCysProThrCysHisGlnAsnAspVal 107
 830 CTGGAGACCCAGATGACAAACAGGTCCACACAGTGTACCAATGATGTTCT 781
 107 erProAspAlaLeuIleAlaAsnLysPheLeuArgGlnAlaValAsn 123
 780 CCCG...ATGCTTTACTGCCACAGTTTGTACGACAGGCTGTGAATTAAC 734
 124 PheLysAsnGluThrGly.TyrThrLysArgLeuArgLysGlnLeuPro 140
 733 TTTAAATAATGAACCTGCTATAC...AGACGACTACGAACACGTTACCTC 687
 140 roPheLeuPheLeuValProProProArgProLeuSerGlnArgAsnLeu 156
 686 CACCCACCCACCCAGTACACACCAACAGCAGTACGACGAGGAGACCTG 637
 157 GlnProArgSerArgSerProIleLeuArgGlnGlnAspProValAlp 173
 636 CAGCCTCTGATGAGATCTCCATATCAAGACAGAGAGATCTCTGATGAT 587
 173 eArgTyThrValSerProThrCysSerAspThrLysThrAlaGlySerC 190
 586 TCCAGTGACATCGCTCCGCTCAGTCA..... 559
 190 ysSerAspSerGlyThrLeuSerArgLeuProAlaProSerIleSerSer 206
 558GTCCTCTATATCTTCA 541
 207 LeuThrSerAsnGlnSerSerLeuAlaProProValSerGlyAsnProSe 223
 540 GTTAACCTTCAATCATCTGCTGCTGCTCTCTCTCTCTGGAATTCATC 491
 223 rSerAlaProAlaProValProAspIleThrAlaThrValSerIleSerV 240
 490 TTTGCCCCCAGCTCCAGTACCTGATATACGCAACCGTGTCTATACAG 441
 240 aHisSerGlyLysSerSerAspGlyProPheArgAspSerAspAsnLysLeu 256
 440 TCCAGTCAGAAAAATCGATGAGACCTTTTGGGATTCGTAATAAATTA 391
 257 LeuProAlaAlaAlaLeuThrSerGluHisSerLysGlyAlaSerSerI 273
 390 TTGGCAGCTGGCGGCTTATCATCGAACAATTCAGAGAGAGAGCTTTCAT 341
 273 eAlaIleThrAlaLeuMetGluLysGlyValAlpProGlyThrSerProt 290
 340 TGTATTTACTGCTCTTATGGAAGAAAGGGGTA.CCAGGTCCAGTCTCT 292
 290 rPAsnSerIlePheVal.GlyGlnSerLeuLeuHisGlyGlnLeuIlePr 306
 291 GGAACCTCATCTTTGTGGACAGTCAATTATTTACATGACAGCTTGATTC 242
 306 oThrThrGlyProValArgIleAsnAlaAlaArgProGlyGlyLysArgP 323
 241 CACAACCTGGCCCACTAAGAAATCATGCTGCTGCTCCAGGTGTGTGGCGGC 192

```

333  roGlyTrpGluHisSerAsnLysLeuGlyTyrLeuValSerProGln 339
|||||
191  CAGCGTGGAGCATTCACACAGCTGGGTACTGATTCTCCACACAG 142
|||||
340  GlnIleuArgArgGlyGluArgSerCysTyrArgSerIleAsnArgGlyAr 356
|||||
141  CAAATTAGAGAGAGAAAGAGCTGTACAGAGCTATTAACCGCGGCG 92
|||||
356  gHisHisSerGluArgSerGlnArgThrGlnSerProSerLeuProAlaT 373
|||||
91  ACACACACAGCAGCATCAGACAGCATCAAGCCCATCATTCCAGCAA 42
|||||
373  hrProCysPheValProValProProPro 382
|||||
41  CTCAGATCTTTGTGCCCCGTTCCACCACT 13
|||||

seq_name: gb_est2:BG575299

seq_documentation_block:
LOCUS      BG575299                820 bp    mRNA    linear    EST 10-APR-2001
DEFINITION  602597953F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4706773 5',
mRNA sequence.
ACCESSION  BG575299
VERSION    BG575299.1 GI:13582952
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 820)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabs-remail.nih.gov
            Tissue Procurement: DCTD/DTF
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.lnl.gov
            Plate: L1M10575 row: 1 column: 14
            High quality sequence stop: 818.
FEATURES             Location/Qualifiers
     source            1..820
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4706773"
                     /clone_1lb="NIH_MGC_87"
                     /tissue_type="mammary adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: Nct1;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.383 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC library."
BASE COUNT      207 a      252 c      155 g      206 t
ORIGIN
alignment_scores:
    Quality: 1073.00      Length: 282
    Ratio: 4.490          Gaps: 5
    Percent Similarity: 84.752      Percent Identity: 80.142

alignment_block:
US-09-811-045A-1 x BG575299
Align seg 1/1 to: BG575299 from: 1 to: 820
134 LeuArgLysGlnLeuProPheLeuPheLeuValProProProArgPr 150

```

```

24  GTCCGAAAACAGTTACCTTCCTCCACCACCCCAATACCACTCCGAGACC 73
|||||
150  oLeuSerGlnArgAsnLeuGlnProArgSerArgSerProIleuArg 167
|||||
74  ACTGATTACAGAGAACCTACCAACCTCTGATGAGATCTCGATATATAAGC 123
|||||
167  IngInasProValAlaPheArgTyrThrValSerProThrCysSerAsp 183
|||||
124  AACAGATATCTCTTATGATTCAGATGACATCTTCATCAACT..... 164
|||||
184  ThrLysThrAlaGlySerCysSerAspSerGlyThrLeuSerArgLeuPr 200
|||||
165  .....CACCC 169
|||||
200  oAlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAlaProP 217
|||||
170  AGCTCCGCTATATCTTCTTACTTCTATACAGTCTTCTTGGCCCTC 219
|||||
217  roValSerGlyAsnProSerSerAlaProAlaProValProAspIleThr 233
|||||
220  CTGTGCTGGAATATCCGCTTCTGCTCCAGCTCCTGATCATATTAAC 269
|||||
234  AlaThrValSerIleSerValHisSerGluLysSerAspGlyProPheAr 250
|||||
270  GCACAGATATCATATCATTCATTCAAGAAAATCAGATGAGACCTTTTCG 319
|||||
250  gAspSerAspAsnLysLeuProAlaAlaAlaLeuHisSerGluHis 267
|||||
320  GGATTCTATATTAATTAATATGCGAGCTGCACCTTCTTCATCAGAGCA 369
|||||
267  erLysGlyAlaSerSerIleAlaIleThrAlaLeuMetGlnLysGly 283
|||||
370  CAAAGGAACCTCCCTCAATTCGATACCGCTTATGGAAGAGAAGGT 419
|||||
284  ValProGlyThrSerProThrAsnSerIlePhe.ValGlyGlnSerLeuL 300
|||||
420  TA.CCAGGTGCTCTTCTTGGAAACCCCAATCTTGGCTGACAGCATATAT 468
|||||
300  euHisGlyGlnLeuIleProThr.ThrGlyProValArgIleAsnAlaAl 316
|||||
469  TGCATGACAGATTGATCCCAACAACTGCTCCAGTAAGAATTAATATATGC 518
|||||
519  TCGTCCAGGTGTGCTGCACACGAGCTGGGAACATTCAACAACCTTGCT 568
|||||
333  yrLeuValSerProGlnGlnIleArgArgGlyGluArgSerCysTyr 349
|||||
569  ATCTGTTTCTCCACACACAAATTAAGAAGGAGGAGAGAGCTGCTAC 618
|||||
350  ArgSerIleAsnArgGlyArgHisHisSerGluArgSerGlnArgThrGl 366
|||||
619  AGAAGTATTAACCGTGGCGACACACAGCAAGAAATATCAAGAGAGCTCA 668
|||||
366  nSerProSerLeuProAlaThrProCysPheValProValProProP 383
|||||
669  AGGCCGGTCACTACACAGCACTCCAGTCTTGTACTTTCACACACT.C 717
|||||
383  roLeuTyrProProPro.ProHisThrLeuProLeuProProGlyValPr 399
|||||
718  CTTTGTATCCGCTTCCTCCCATACACTTCTCTCTCCGGGGGTGCTC 767
|||||
399  o.ProProGlnPheSerProGlnPheProSerSerGln 411
|||||
768  TCCTCCACAGTTTCTCTCAGTTTCTCTCTGCGCAG 805
|||||

seq_name: gb_est2:BG622049

seq_documentation_block:
LOCUS      BG622049                847 bp    mRNA    linear    EST 18-APR-2001
DEFINITION  602646655F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4768333 5',
mRNA sequence.

```


700 rplySerrLySgluLySgluSerGluAsnValProGluLySpgLyLySgLy 716
|||||
253 CAAAGTCAAAAGAGAGAGAGTGAAGAACGCTCCAGAGATGGTAAAGGA 302
717 AsnLyShLySlySlyHisArGlyArGTAAGAsnGluGluLySgLyGluGlu 733
|||||
303 AATAAGCATTAAGAACACAGAAAGACAAAAA.....AAGGGAGGA 343
733 uSerGluSerPheLeuAsnProGluLeuLeuGluThrSerArGlyScySA 750
|||||
344 AAGTGAAGGGTTTTCGAACCCAGAGATTATTAGAGACTTCTAGAAATCAA 393
750 rGgLySerSerGlyIleAspGluThrLySThrAspThrLeuPheValLeu 766
|||
394 GAGAACCTCAGAGTGTGGAAGAAATAAAAACAGACTCAATTGTTGTCYC 443
767 ProSerArGAspAspAlaThrProValArgAspGluProMetAspIleGlu 783
|||||
444 CCAAGAGAGATGATGCCACACCTGTTAGAGATGAACCATGGATGATCCAGA 493
783 uSerIleThrPheLySserValSerAspLyAspLySArgGluLySAsp 799
|||||
494 ATCATTCACCTTTTAATCATCGTGTCTGAAGAAAGACAGAGAGAAAGCAT 543
800 LySProLySValLySserAspLySThrLySArGlySerAspGlySerAl 816
|||||
544 AAACCAAAAGCAAGGGTGATTAACCAACGAGAGAAATGATGATCTGC 593
816 aThrAlaLySlySAspAsnValLeuLySProSerLySgLyProGluGlu 832
594 TGTGTCCAAAAAAGAC...ATATGTGAACACTGCTAAAGAGACCCCAAG 640
832 uLySValAspGlyAspArgGluLyS...SerProArgSerGluProProle 848
|||||
641 AAAAGATAGATGGAGAACGTGAGAGATTCTCTCGATTCTGAACCTCCAT 690
848 u...LySlySAlaLySgLuGluAlaThrLySlyleAspSerValLySPro 864
|||
691 TTAAGAGAAAGCCAAAGAGGAGATCCGAGAGACTGACATTAATTAATCAT 740
864 eSerSerSerGluLySAsp...GluLySValThrGlyThrProArLyS 880
|||||
741 CATCTTCCTTCAGAAAGATGAAAAATTCCTCTGTAACCCCAAGAAAG 790
880 laHisSerLySerAlaLySAspThrArGArgGlnSerGln 893
|||||
791 TTCATCTTCTTAATCAGAGCAAAAGACCAACAAGAACAAACACAG 893
seq_name: gb_est2:BI333743
seq_documentation_block: 771 bp mRNA linear EST 30-JUL-2001
LOCUS BI333743
DEFINITION 60299921.F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5141261 5',
mRNA sequence.
ACCESSION BI333743
VERSION BI333743.1 GI:15018400
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 771)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@p5-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LINL at:
 http://image.lnl.gov
 Plate: L1AM1348 row: a column: 06
 High quality sequence stop: 768.

FEATURES

source
 1. 771
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="5141261"
 /clone_id="NIH_MGC_12"
 /tissue_type="cervical carcinoma cell line"
 /lab_host="DH10B"
 /note="Organ: cervix; Vector: PCMV-SF0RT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.4 kb. Library prepared by Life
 Technologies."
 BASE COUNT 249 a 181 c 136 g 205 t
 ORIGIN

alignment_scores:
 Quality: 1044.00 Length: 248
 Ratio: 4.879 Gaps: 2
 Percent Similarity: 86.290 Percent Identity: 83.871

alignment_block:
 US-09-811-045A-1 x B1333743 ..

Align seg 1/1 to: B1333743 from: 1 to: 771

```

1 MetMetGluValLysAspProAsnMetLysGlyAlaMetLeuThrAsnTh 17
83 ATGATGAGGAGTGAAGATCCATATGAAAGGTGCAATCCTTACCAACAC 132
17 TGLYSTRYAlaIleProThrIleAspAlaGluAlaTyrAlaIleGlyL 34
133 TGGAAATATGCAATACCACTATAGATGACAGACATATGCAATTTGGGA 182
34 YSLYSGluLysProProPheLeuProGluGluProSerSerSerSerGlu 50
183 AGAAGAGAAACCTCCCTTCTTACCAAGAGACCATCTTCTTCCACAGA 232
51 GUAASPAPProIleProAlaGluLeuLeuCysLeuIleCysLysAspTl 67
233 GAAGATGATCCATATCCAGATGAATGTTGTCTCATCTGCACAGATAT 282
67 GetherAspAlaValIleProCysCysGlyLysAsnSerSerCysAspG 84
283 TATGACTGATGCTGTGTGATTCCTGCTGCTGGAACGTTTACTGTGATG 332
84 LuCysIleLeuThrLeuLeuGluSerAspLysHisThrCysProThr 100
333 AATGATATAGAACACACTCCTCGAATCAGATGACACACATGTCGAGCG 382
101 CysHisGlnAsnAspValSerProAspAlaLeuIleAlaAsnLysPheLe 117
383 TGTCTCAAAAAGATGTTCTCCTCATGCTTAAATGGCCAAATTAATTTT 432
117 uArgGlnAlaValAsnAspPheLysAsnGluThrGlyTyrThrLysArgL 134
433 ACGACAGGCTGTAATAACTTCAAAAATGAACTGCTTATACAAAAGAC 482
134 euArgLysGlnLeuProProPheLeuPheLeuValProProArgPro 150
483 TACGAAAACAGTTACTCTCTCCACACCCCAATACCACTCCGAGACCA 532
151 LeuSerGlnArgAsnLeuGlnProArgSerArgSerProIleLeuArgL 167
533 CTGATTCAGAGAACCTACCAACCTCGATGAGATCTCCGATATCAAGACA 582
167 nGlnArgProValIlePheArgTyrThrValSerProThrCysSerAspT 184
583 ACAAGATCCTTATGATTCACAGTACATCTTCAATCAACT..... 622

```

```

184 hrLysThrAlaGlySerCysSerAspSerGlyThrLeuSerArgLeuPro 200
623 .....CACCA 628
201 AlaProSerIleSerSerLeuThrSerAsnGlnSer.SerLeuAlaProP 217
629 GCTCGCTATATCTTCAATTAATCAATGTTCTTCCCTTGGCCCTGC 678
217 rovalSerGlyAsnProSerSerSerAlaProAlaProValProAspLeuThr 233
679 CTGCTCTGGAATATCCGTTCTTCTGCTCCAGCTCCTGTTACCTATTA 728
234 AlaThrValSerIleSerValHisSerGluLysSerAspGly 247
729 GCAACAGTATCCATATCAATTCATTCAGAAAATATCAGATGCA 770

```

seq_name: gb_est2:BF720766

seq_documentation_block:

LOCUS BF720766 744 bp mRNA linear EST 03-JAN-2001
 DEFINITION mab58d02.y1 Soares.thymus.2NDMT Mus musculus cDNA clone
 IMAGE:3974450 5' similar to TR:P70287 P70287 RETINOBLASTOMA BINDING
 PROTEIN 6 ; mRNA sequence.

ACCESSION BF720766
 VERSION BF720766.1 GI:12021768
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 TITLE (bases 1 to 744)
 NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

COMMENT
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 This clone is available royalty-free through LINL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:1474482

Seq primer: -40RP from Gibco
 High quality sequence stop: 478.

FEATURES

source
 1. 744
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_image="3974450"
 /clone_id="Soares_thymus_2NDMT"
 /sex="male"
 /tissue_type="Thymus"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - Oligo(dT) primer [5',
 TGTATCAATCTGAAGTGGAGCGCGCGGTTTTTTTTTTTTTTTTTTTT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. RNA
 provided by Dr. Bertrand Jordan. Library went through two
 rounds of normalization, and was constructed by Bento
 Soares and M.Fatima Bonaldo."
 BASE COUNT 187 a 233 c 138 g 180 t 6 others
 ORIGIN

alignment_scores:
 Quality: 1042.00 Length: 267
 Ratio: 4.673 Gaps: 2
 Percent Similarity: 83.521 Percent Identity: 80.524


```

861 llysProSerSerSerGlnLysaspGluLysValThrGlyThrPro 878
      |||
201 AAAACCTCTCTCTCTCTCAGAGATGAGAAAGTACAGAACCCCTA 250
      |||
878 rglYsalahISerLysSerAlaLysasp-ThrArgArgGlnSerGlnPr 894
      |||
251 GAAACCCCTCTCTAAATCTGCAAAAGAACACGAGGAGCAAGCCGCC 300
      |||
894 oArgThrArgSerSerLysArgThrValProLysThrSerSerGlnLys 911
      |||
301 AAGGAGAGAGAGGTCAAAAAGAGCTGTTCCAAAGACATCAAGTCAAGAA 350
      |||
911 exGlnProValArgThrArgArgProArgSerLeuArgLysIleasnTy 927
      |||
351 GCCAGCAGTAAGAGAGAGAGGCAAGAGCCTGAGAAAAATAAATAC 400
      |||
928 leuIleAlaArgGluLysasnGluArgGluLysArgLys..LysSerVal 943
      |||
401 TTGATAGCAAGGAGGAGAAAGCAAGAGGAGAAAGGAGAAAGAGTGT 450
      |||
944 AspLysAspPheGluSerSerSerMetLysIleSerLysValGluGlyTh 960
      |||
451 GATTAAGATTTTGTAGTCGTCTTCAATGAAATCTTAAGTAAAGAGAAC 500
      |||
960 rgluIleValLysProSerProLysArgLysMetGluGlyAspValGlu 977
      |||
501 AGAAATAGTGAACCATCACTGAAATGGAAGAGTGTGTGAA 550
      |||
977 ystLeuLysArgThrProGluLysAspLysIleAlaSerSerThrThrPro 993
      |||
551 AGCTGGAAGAGACCCAGTAAGAGACAAAGTTCATCATCACTACTCTCA 600
      |||
994 AlaLysLysIleLysLeuasnArgGluThrGlyLysLysIleGlyAsnAl 1010
      |||
601 GCCAAAAAATCAAACTTCACAGAGAAACGCAAAAAAATTTGGAATGC 650
      |||
1010 acGluAsnAlaSerThrThrLysGluPro 1019
      |||
651 AGAAATGCATCTGTAGTCAAGAACCCCT 678
seq_name: gb_gss:A2696947

seq_documentation_block:
LOCUS      A2696947              782 bp      DNA      linear      GSS 24-JAN-2001
DEFINITION RPCI-23-238D2.TJ RPCI-23 Mus musculus genomic clone RPCI-23-238D2,
DNA sequence.
ACCESSION  A2696947
VERSION    A2696947.1 GI:12412563
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 782)
            Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akintret
            ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
            Mouse BAC End Sequences from Library RPCI-23
            Unpublished (1999)
            Other GSS: RPCI-23-238D2.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhaog@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pjejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 238 row: D column: 2

```

```

Seq primer: SP6
Class: BAC ends.
FEATURES             Location/Qualifiers
     source            1..782
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-238D2"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site-1:
                     EcoRI; Site-2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and partially digested
                     with a combination of EcoRI and EcoRI Methylase. Size
                     selected DNA was cloned into the pBAC3.6 vector at the
                     EcoRI sites. The ligation products were transformed into
                     DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      186 a      145 c      249 g      202 t
ORIGIN
alignment_scores:
      quality: 995.00      length: 255
      ratio: 4.502      gaps: 1
Percent similarity: 86.667      Percent identity: 76.863
alignment_block:
US-09-811-045a-1 x A2696947/rev ..
Align seg 1/1 to reverse of: A2696947 from: 1 to: 782
200 ProAlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAlaPr 216
      |||
762 CCACACCTCCCTGTACACCTATACCTTCAATTCATCTTCCGCGCTC 713
      |||
216 opProValSerGlyAsnProSerSerAlaProAlaProAlaProAlaPr 233
      |||
712 TTCTGTGCGCTGGAACCCATCTCTCCCGAGCTCCGCTCCGAGAACCA 663
      |||
233 hvrAlaThrValSerIleSerValHisSerGluLysSerAspGlyProPhe 249
      |||
662 CTCGACAGAGTCTGTATCATGACCTTTCAGAAAAACAGATGATCTT 613
      |||
250 ArgAspSerAspAsnLysLeuLeuProAlaAlaAlaLeuThrSerGlnH 266
      |||
612 CGCGATTCAGAAACCAATATTATCACGATCGCCGCTTACATCAGAAC 563
      |||
266 sSerLysGlyAlaSerSerIleAlaIleThrAlaLeuMetGluLysG 283
      |||
562 TTCAAAGGAGGCGCTTCAATTCGTGTACTCTGCTCTGATGAGAAAGC 513
      |||
283 lylAlaProGlyThrSerProThrPasnSerIlePheAlaGlyInSerie 299
      |||
512 GTGGC.CCAGGTGGCGGTCCTGAAACTCCACCTTCGTGGACAGTCAC 464
      |||
299 uLeuHisGlyGlnLeuIleleProThrThrGlyProValArgLysAsnAl 316
      |||
463 ACTCCATGAACAGTCTATCCCACTGCTGTCAGTAAGATTAAGTCG 414
      |||
316 laArgProGlyGlyArgProGlyTyrGlnHisSerAsnLysLeuGly 332
      |||
413 CTCATCCAGTGGTGTCAACAGGTTGGAAACATTCACAAAGATVGG 364
      |||
333 TyrLeuValSerProProGlnGlnIleArgArgGlyGluArgSerTy 349
      |||
363 AAGCTAGTTTCTCAACACAGCAATTTAGAGTGGAAGAAAGAGCTGTA 314
      |||
349 rArgSerIleAsnArgGlyArgHisSerSerGluArgSerGlnArgTh 366
      |||
313 CCAGAGATTAACAGTGGCGAGACTACAGTGAACATCATCAGAGAGAG 264
      |||
366 lserProSerLeuProAlaThrProCysPheValProValProProPro 382

```

[illegible]

PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2002, 16:19:12 ; Search time 69.71 Seconds

(without alignments)
3484.219 Million cell updates/sec

Title: US-09-811-045A-1
Perfect score: 7374
Sequence: 1 MMEVXDPNMGAMLTWTGKT.....WRSWLLGRWRAPSSRNPS 1404

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP_archaea:*
2: SP_bacteria:*
3: SP_fungi:*
4: SP_human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_prodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*
15: SP_virus:*
16: SP_bacteriophage:*
17: SP_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6054	82.1	1591	11	P97868
2	5451.5	73.9	1616	4	Q96PH3
3	3500.5	47.5	948	4	Q15290
4	1702	23.1	628	4	Q9H5M5
5	1305.5	17.7	529	4	Q9NFX4
6	542	7.4	1231	5	Q9X221
7	480.5	6.5	1190	5	Q62235
8	405	5.5	2081	10	Q9LH98
9	396.5	5.4	2752	4	Q9UQ35
10	396	5.4	4880	11	Q9UL71
11	396	5.4	5085	11	Q9UKS6
12	388.5	5.3	5120	13	Q9PU36
13	384	5.2	3484	5	P91257
14	377.5	5.1	2296	5	Q9UHA2
15	375	5.1	17352	5	Q95YM2
16	374.5	5.1	5327	5	Q76891

17	367.5	5.0	4833	11	Q9QYX6	Q9QYX6 mus musculu
18	367.5	5.0	5038	11	Q9QYX7	Q9QYX7 mus musculu
19	363.5	4.9	6632	5	Q17362	Q17362 caenorhabdi
20	363.5	4.9	6642	5	Q01761	Q01761 caenorhabdi
21	360.5	4.9	2083	5	Q9N435	Q9N435 caenorhabdi
22	351	4.8	2768	5	Q9VC00	Q9VC00 drosophila
23	349	4.7	990	13	Q91803	Q91803 xenopus lae
24	344.5	4.7	3111	5	Q9VH10	Q9VH10 drosophila
25	341.5	4.6	1422	6	Q9SKU4	Q9SKU4 canis famli
26	340	4.6	5322	5	Q9VPE9	Q9VPE9 drosophila
27	340	4.6	5322	5	Q9VPE9	Q9VPE9 drosophila
28	339	4.6	2722	5	Q9VPE9	Q9VPE9 drosophila
29	338.5	4.6	2157	11	Q9Z1R1	Q9Z1R1 mus musculu
30	336.5	4.6	4498	13	Q93291	Q93291 fuqu rubrip
31	336	4.6	1062	5	Q960C4	Q960C4 drosophila
32	335.5	4.5	933	4	Q9UQ14	Q9UQ14 homo sapien
33	335.5	4.5	1026	4	Q9UJ57	Q9UJ57 homo sapien
34	333	4.5	1108	5	Q9ND10	Q9ND10 babesia big
35	333	4.5	1262	4	Q9UQ40	Q9UQ40 homo sapien
36	333	4.5	1343	4	Q9H7N4	Q9H7N4 homo sapien
37	332.5	4.5	1151	5	Q9VAY4	Q9VAY4 drosophila
38	331	4.5	2897	5	Q9VID9	Q9VID9 drosophila
39	331	4.5	2951	5	Q9W320	Q9W320 drosophila
40	330	4.5	1430	5	Q9W0B0	Q9W0B0 drosophila
41	329	4.5	1160	11	Q9ESC8	Q9ESC8 mus musculu
42	328	4.4	2701	4	Q9Y520	Q9Y520 homo sapien
43	327	4.4	2187	11	P70670	P70670 mus musculu
44	326.5	4.4	1341	4	Q9UKV3	Q9UKV3 homo sapien
45	325.5	4.4	1178	5	Q9NHX5	Q9NHX5 drosophila

ALIGNMENTS

RESULT 1
P97868 PRELIMINARY; PRT; 1591 AA.
ID P97868; P70287;
AC P97868; P70287;
DT 01-MAY-1997 (TREMBLREL. 03. Created)
DT 01-OCT-2000 (TREMBLREL. 15. Last sequence update)
DT 01-JUN-2001 (TREMBLREL. 17. Last annotation update)
DE RETINOBLASTOMA BINDING PROTEIN 6 (PROLIFERATION POTENTIAL-RELATED
DE PROTEIN) (P53-ASSOCIATED CELLULAR PROTEIN) (PACT).
GN RBBP6 OR P2P-R.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY (ISOFORM 2).
RC STRAIN-BALB/C;
RX MEDLINE-97188447; PubMed-9037032;
RA Witte M.M., Scott R.E.;
RT "The proliferation potential protein-related (P2P-R) gene with domains
RT encoding heterogeneous nuclear ribonucleoprotein association and Rb1
RT binding shows repressed expression during terminal differentiation.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:1212-1217(1997).
RN [2]
RP REVISIONS.
RC STRAIN-BALB/C;
RX Witte M.M., Scott R.E.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 4-1587 FROM N.A., AND CHARACTERIZATION (ISOFORMS 1 AND 2).
RC STRAIN-BALB/C; TISSUE-TESTIS;
RX MEDLINE-97163396; PubMed-9010216;
RA Simons A., Melamed-Bessudo C., Wolkowicz R., Sperling J., Sperling R.,
RA Eisenbach L., Rotter V.;
RT "PACT: Cloning and characterization of a cellular p53 binding protein
RT that interacts with Rb1";
RL Oncogene 14:145-155(1997).
CC -!- FUNCTION: BINDS TO Rb1 THROUGH THE Rb1 POCKET DOMAIN. ASSOCIATES
CC WITH HNRNP. ALSO BINDS TO P53. MAY BE INVOLVED IN GROWTH CONTROL

CC AND DIFFERENTIATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS: ISOFORM 1 (SHOWN HERE) AND
CC ISOFORM 2: MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN TESTIS.
CC MODERATE LEVELS FOUND IN HEART, LUNG, LIVER AND SKELETAL MUSCLE.
CC VERY LOW LEVELS DETECTED IN KIDNEY, BRAIN AND SPLEEN. ALSO
CC EXPRESSED IN GROWING UNDIFFERENTIATED MESECHYMAL STEM CELLS BUT
CC NOT DURING THE TERMINAL DIFFERENTIATION PHASE.
CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR EMBL: U83913; AAC72432.1; -
DR EMBL: U28789; AAB49620.1; -
DR MGI: 894835; Rdbp6.
DR InterPro: IPR001841; Znf_fing.
DR Pfam: PF00097; zf-C3HC4; 1.
DR SMART: SM00184; RING: 1.
DR PROSITE: PS00518; ZINC_FINGER_C3HC4; FALSE_NEG.
KW Zinc-finger: Metal-binding; DNA-binding; Nuclear protein;
KW Alternative splicing.
FT ZN_FING 61 101
FT DOMAIN 481 540
FT VARSLIC 1528 1585
FT CONFLICT 454 487
FT CONFLICT 55 55
FT CONFLICT 118 119
FT CONFLICT 142 142
FT CONFLICT 219 219
FT CONFLICT 222 222
FT CONFLICT 381 381
FT CONFLICT 396 396
FT CONFLICT 416 423
FT CONFLICT 430 430
FT CONFLICT 437 437
FT CONFLICT 448 448
FT CONFLICT 490 490
FT CONFLICT 504 504
FT CONFLICT 590 590
FT CONFLICT 742 742
FT CONFLICT 757 759
FT CONFLICT 764 764
FT CONFLICT 779 779
FT CONFLICT 783 783
FT CONFLICT 813 813
FT CONFLICT 1091 1091
FT CONFLICT 1117 1117
FT CONFLICT 1365 1365
FT CONFLICT 1382 1382
FT CONFLICT 1392 1392
FT CONFLICT 1397 1397
FT CONFLICT 1405 1405
SQ SEQUENCE 1591 AA; 177763 MW; 7D0C6798BF49D3748 CRC64;

Query Match 82.1%; Score 6054; DB 11; Length 1591;
Best Local Similarity 84.7%; Pred. No. 1.3e-309;
Matches 1199; Conservative 31; Mismatches 111; Indels 74; Gaps 10;

OY 1 MAEVDPPNKGMLNTGKYAIPITDAEYVAGKKEKPPLEPESSSEEDDPIPAELL 60
DB 1 MAEVDPPNKGMLNTGKYAIPITDAEYVAGKKEKPPLEPESSSEEDDPIPAELL 60
OY 61 CLICDINTDAVAVICCCNSSCDECIRFTLLSDKHCPTCHQNDVSPALANKFLQA 120
DB 61 CLICDINTDAVAVICCCNSSCDECIRFTLLSDKHCPTCHQNDVSPALANKFLQA 120
OY 121 VNNFKNETGYTKRLKQPLFLVPPRPRLSQRNLOPRSSPILRODDPVVFRVTSPT 180
DB 121 VNNFKNETGYTKRLKQPLFLVPPRPRLSQRNLOPRSSPILRODDPVVFRVTSPT 180
OY 181 CSDFKTAGSCSDSGLSLRLPAPSSISLTSNOSLAPVSGSPSAPAPVDPITATVISIV 240
DB 181 CSDFKTAGSCSDSGLSLRLPAPSSISLTSNOSLAPVSGSPSAPAPVDPITATVISIV 240
OY 181 HS-----APSSISLTSNOSLAPVSGSPSAPAPVDPITATVISIV 222

OY 241 HSEKSDGPPRDSNKLPLAALTSEHSKASSIATATLMEKC-----VQSPWMSIFVG 296
DB 223 HSEKSDGPPRDSNKLPLAALTSEHSKASSIATATLMEKC-----VQSPWMSIFVG 296
OY 297 QSLHQLIPTTGYPVIRINARPGGRGWHENKLGTVSPPOQIRGRSCYRSTINCR 356
DB 279 QSLHQLIPTTGYPVIRINARPGGRGWHENKLGTVSPPOQIRGRSCYRSTINCR 356
OY 357 HSEKSORPSLPLATCFVVPVPPPLYPVPPHPLPLPPGVPQSPQFSSQPTAG 416
DB 339 HSEKSORPSLPLATCFVVPVPPPLYPVPPHPLPLPPGVPQSPQFSSQPTAG 416
OY 417 YSVPPGPPAPANISTACFSQVPTAHSNMPTQAOLSLREBFYRQ----- 465
DB 399 YSVPPGPPAPANISTACFSQVPTAHSNMPTQAOLSLREBFYRQ----- 465
OY 466 -----NDKGE-----SKFPYSGSSSYSTDS-SQGLAHIALT 501
DB 459 KLDEFTDAFKELMEYKKIKERRRSRSPSSGSSYSSSTYSKSSGSTRSRSS 518
OY 502 LSPSAHTLTLHDHP-----PEAFARSAMIVHMDLMDIAHARSPPYRRRSR 557
DB 519 RSPSRSHSRYSRSPYPRGRGKSRNYSRSH-----GYHRSRSRSPYRRRSR 573
OY 558 SPPEFGQSPTRKNVREKEREYFNRYRPPYDIKAYGSDVPDPFEKERYRME 617
DB 574 SPQAFGQSPTRKNVREKEREYFNRYRPPYDIKAYGSDVPDPFEKERYRME 617
OY 618 RYREYKERYGYAVGAOPRPSANREDSPELPLNTRNSPFTGRREDYAAQOSH 677
DB 634 RYREYKERYGYAVGAOPRPSANREDSPELPLNTRNSPFTGRREDYAAQOSH 677
OY 678 RNLGNYPEKLTROSHNANKDPKSKESSEVPPDGGKGNKHKHRRKNEKSESE 693
DB 694 RNLGNYPEKLTROSHNANKDPKSKESSEVPPDGGKGNKHKHRRKNEKSESE 693
OY 738 LNPelletSRKCRGSGGIDETKTDTLFLVPSRDATPVPRDPMNAESTTFY 797
DB 751 LNPelletSRKCRGSGGIDETKTDTLFLVPSRDATPVPRDPMNAESTTFY 797
OY 798 KDKPKVKSDDTKRRSDGATAKKDNVLPKSGPOEKYDGDREKSPREPLKAKKE 857
DB 811 KDKPKVKSDDTKRRSDGATAKKDNVLPKSGPOEKYDGDREKSPREPLKAKKE 857
OY 858 IDSVKPSSSSKDEKVTGTPKAKSKSAKDRROSOQRTKSKTVPKTSOSQOPY 917
DB 871 IDSVKPSSSSKDEKVTGTPKAKSKSAKDRROSOQRTKSKTVPKTSOSQOPY 917
OY 918 RPSLRKINYLILAR-EKNEREKRRKSYDXDFESSSMKISKVEGETIYKPPK 930
DB 931 KAKKPEKNKLLDSKGEKRRKRTKESYDQFESSSMKISKVEGETIYKPPK 930
OY 977 KLERTEKDIASSTPAKKIKINRETGKKGNAENASTTKREPSEKLESTSSK 1036
DB 991 KLERTEKDIASSTPAKKIKINRETGKKGNAENASTTKREPSEKLESTSSK 1036
OY 1037 GAKKRVVAGSESSSTLVDTYSTSGSGSPVRSKSEKTKTKTVIKMEYVND 1096
DB 1051 GAKKRVVAGSESSSTLVDTYSTSGSGSPVRSKSEKTKTKTVIKMEYVND 1096
OY 1097 DVYIMIOVQSKWMDDESEEDVKTQPIOSVKGPSIINKVTKSATAKATY 1110
DB 1111 DVYIMIOVQSKWMDDESEEDVKTQPIOSVKGPSIINKVTKSATAKATY 1110
OY 1157 QPEKLOKLPKASHLMOHELRSKSGASSSEKGRAKDRHSGSEKDNDRKSGAOP 1216
DB 1171 QPEKLOKLPKASHLMOHELRSKSGASSSEKGRAKDRHSGSEKDNDRKSGAOP 1216
OY 1217 STVDRLSEOGHFKTLSSSKETRTSEKHESEYRGSSNDDFTPGRKVDYDS 1276
DB 1231 STVDRLSEOGHFKTLSSSKETRTSEKHESEYRGSSNDDFTPGRKVDYDS 1276
OY 1277 DERGELARRKDSPPRGKESLSGOKSKLREEDLPPKGAESKSKNSPPRDKK 1336

DB 1291 DEGGELARRKDSPPKRESLSGOKSLREERDLPKKAESKSNSSPPDKKHHDKAPY 1350
 OY 1337 ETKRCEETKPYDKNSGKERKHAHAEARNKESG 1371
 DB 1351 ETKRCEETKPYDKNSGKERKHAHAEARNKESG 1385

RESULT 2
 OY 096PH3 PRELIMINARY: PRT: 1616 AA.

AC 096PH3: 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE PROLIFERATION POTENTIAL-RELATED PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Camargo A.A., Moreira E.S., Simpson A.J.G.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF352051; AAL05625.1;
 SQ SEQUENCE 1616 AA: 182248 MW: D89F5B4BED95D72B CRC64:

Query Match 73.9%; Score 5451.5; DB 4; Length 1616;
 Best Local Similarity 75.3%; Pred. No. 5.6e-278;
 Matches 1073; Conservative 92; Mismatches 167; Indels 93; Gaps 16;

OY 1 MMEVDPMMKAMLTNTGKVAIPTIDAEVAIGKKEKPPLEPSSSEEDDPIPAELL 60
 DB 23 MMEVDPMMKAMLTNTGKVAIPTIDAEVAIGKKEKPPLEPSSSEEDDPIPAELL 82
 OY 61 CLICKDITDAVAVIPCCGNSCDECIRTTTLESKHTCTCHQNDVSPDALIANKELROA 120
 DB 83 CLICKDITDAVAVIPCCGNSCDECIRTTTLESKHTCTCHQNDVSPDALIANKELROA 142
 OY 121 VNNFNENGYKTKRLKQKLPFLVPPRPRLSORNLQPRSRPIRQODPVVFTVSP 180
 DB 143 VNNFNENGYKTKRLKQKLPFLVPPRPRLSORNLQPRSRPIRQODPVVFTVSP 202
 OY 181 CSDTAGSCSDSGTSLRPAVPAISLTSNOSSLAPVSGNPSAPAPVDTITATVSI 240
 DB 203 CSDTAGSCSDSGTSLRPAVPAISLTSNOSSLAPVSGNPSAPAPVDTITATVSI 244
 OY 241 HSEKSDGPPROSDNKLPAALTLSEHSKASIAITALEENG----VPGTSPMNSIPV 296
 DB 245 HSEKSDGPPROSDNKLPAALTLSEHSKASIAITALEENG----VPGTSPMNSIPV 300
 OY 297 QLLHGOQLPTTGPVIRINARPGGPRGHEHNSKLGVLVSPQQLRGRSCYRSINRGR 356
 DB 301 QLLHGOQLPTTGPVIRINARPGGPRGHEHNSKLGVLVSPQQLRGRSCYRSINRGR 360
 OY 357 HHSERSORTQSPSLPAPCPVPPRPPLPPPHLPLRPGVPPQSPQPPSSQPPAG 416
 DB 361 HHSERSORTQSPSLPAPCPVPPRPPLPPPHLPLRPGVPPQSPQPPSSQPPAG 420
 OY 417 YSVPPGPPAPANISTACSPGVPRAHNTMPTDAPLRSREERFREQ----- 465
 DB 421 YSVPPGPPAPANISTACSPGVPRAHNTMPTDAPLRSREERFREQ----- 480
 OY 466 NDGKRE-----SKPPYSGSSYRSSTYDS--SQGLAQHIAHLT 501
 DB 481 KLDEFNTDEKELMEYKTKIOKERRRFSNKSSTYSSTYSSTYSSTYSSTYSST 540
 OY 502 LSPSAHTLDLHDHPH-----PEAEARSAMIVHMPDLMDIAHARSPPRYRSR 557
 DB 541 RSPSRHSRYSRSPYPRGRGRSRNYSRSH-----GYHRSRSPRYRSR 595
 OY 558 SPPFNGQSPTKRNVPREREKREYFNRYRPPYDILKATYGSVDFRDFEKERTREME 617

DB 596 SPOAFRGQSPNKRNVPOGETEREYFNRYRPPYDMKAYGSGVDFRDFEKERTREME 655
 OY 618 RRYREMYEYKYGYAAGAPRPSANREDFSPERLPLNTRNSPFTGRREDYAAAGSHRN 677
 DB 656 RRYREMYEYKYGYAAGAPRPSANREDFSPERLPLNTRNSPFTGRREDYAAAGSHRN 715
 OY 678 RNLGNYPEKLTSTRDSHNAKNKPNKSKESSENVPGDGKGNKRRKRRNEKEGESESE 737
 DB 716 RNLGNYPEKLTSTRDSHNAKNKPNKSKESSENVPGDGKGNKRRKRRNEKEGESESE 772
 OY 738 LNPLETSTRDSHNAKNKPNKSKESSENVPGDGKGNKRRKRRNEKEGESESE 797
 DB 773 LNPLETSTRDSHNAKNKPNKSKESSENVPGDGKGNKRRKRRNEKEGESESE 832
 OY 798 KOKPYKSDTKRKSOGSATAKKNVLPKSGOEKVDGKREKSPREPLKAKKAKETPK 857
 DB 833 KOKPYKSDTKRKSOGSATAKKNVLPKSGOEKVDGKREKSPREPLKAKKAKETPK 892
 OY 858 IDSVKPSSSQKDEKVTGTPRKAHSAKADTRROSOPTRRSRKRYVKTSSQSPYRTR 917
 DB 893 IDSVKPSSSQKDEKVTGTPRKAHSAKADTRROSOPTRRSRKRYVKTSSQSPYRTR 950
 OY 918 RPRSLRKINYLAREKN-----EREKR---KSVKDESSSMKISKVGTETIYKPS 967
 DB 951 RPRSLRKINYLAREKN-----EREKR---KSVKDESSSMKISKVGTETIYKPS 1004
 OY 968 KRMKEGVKLETPPEKDKIASSSTTPPAKKIKLNENCKKIGNMASTTEPSEKLEST 1027
 DB 1005 KRMKEGVKLETPPEKDKIASSSTTPPAKKIKLNENCKKIGNMASTTEPSEKLEST 1063
 OY 1028 SKIOEKYKGAKRKRVAGSESSSTLVDTSTSTGSGPVAKSEKTDKRTVKTMEY 1087
 DB 1064 SKIOEKYKGAKRKRVAGSESSSTLVDTSTSTGSGPVAKSEKTDKRTVKTMEY 1123
 OY 1088 NNDTAPAEVYIMIQVPOSKWMDKDFESEBEDYKTTPIOSVAKPSSIIINVTTKSAT 1147
 DB 1124 NNDTAPAEVYIMIQVPOSKWMDKDFESEBEDYKTTPIOSVAKPSSIIINVTTKSAT 1183
 OY 1148 AKYTEKESQPEKLOKLPKESHLMQHELRSSGASSEKGRADRHSSEKDNPKR 1207
 DB 1184 AKYTEKESQPEKLOKLPKESHLMQHELRSSGASSEKGRADRHSSEKDNPKR 1243
 OY 1208 KSGAPDSESTYVDRSEOGHFTLSQSKETRTSEKHSYVGSNKDTPGRDKKVDYS 1267
 DB 1244 KSGAPDSESTYVDRSEOGHFTLSQSKETRTSEKHSYVGSNKDTPGRDKKVDYS 1303
 OY 1268 RDYSSSKRDERGELARRKDSPPRKGESLSGOKSLREERDLPKKG-AESKSNSSPPRD 1326
 DB 1304 RDYSSSKRDERGELARRKDSPPRKGESLSGOKSLREERDLPKKG-AESKSNSSPPRD 1363
 OY 1327 KRPDHKAPYETKRCCEETKPYDKNSGKERKHAHAEARNKESG 1371
 DB 1364 KRPDHKAPYETKRCCEETKPYDKNSGKERKHAHAEARNKESG 1408

RESULT 3
 ID 015290 PRELIMINARY: PRT: 948 AA.
 AC 015290:
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE RB PROTEIN BINDING PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-96129310; PubMed-8595913;
 Sakai Y., Saijo M., Coelho K., Kishino T., Niihawa N., Taya Y.;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2002, 16:19:12 : Search time 69.71 Seconds
(without alignments)
3484.219 Million cell updates/sec

Title: US-09-811-045a-1
Perfect score: 7374
Sequence: 1 MAEVEKDPNKGAMLTFTGKY.....WRRSWLLGWRWRAPSSRNPS 1404

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTRMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6054	82.1	1591	11	P97868 mus musculu
2	5451.5	73.9	1616	4	Q96PH3 Q96ph3 homo sapien
3	3500.5	47.5	948	4	Q15290 Q15290 homo sapien
4	1702	23.1	628	4	Q9H5M5 Q9H5m5 homo sapien
5	1305.5	17.7	529	4	Q9NPX4 Q9npX4 homo sapien
6	542	7.4	1231	5	Q9XZ21 Q9xZ21 drosophila
7	480.5	6.5	1190	5	Q62235 Q62235 caenorhabdi
8	405	5.5	2081	10	Q9LH98 Q9Lh98 arabidopsis
9	396.5	5.4	2752	4	Q9U035 Q9u035 homo sapien
10	396	5.4	4880	11	Q9JL71 Q9jL71 rattus norv
11	388.5	5.3	5120	13	Q9PU36 Q9pu36 gallus gall
12	384	5.2	3484	5	P91257 P91257 caenorhabdi
13	377.5	5.1	2296	4	Q9UHAB Q9uhAB homo sapien
14	375	5.1	17352	5	Q95YM2 Q95ym2 procambarus
15	374.5	5.1	5327	5	Q76891 Q76891 drosophila

17	367.5	5.0	4833	11	Q9QYX6 Q9qYx6 mus musculu
18	367.5	5.0	5038	11	Q9QYX7 Q9qYx7 mus musculu
19	363.5	4.9	6632	5	Q17362 Q17362 caenorhabdi
20	363.5	4.9	6642	5	Q01761 Q01761 caenorhabdi
21	360.5	4.8	2083	5	Q9N435 Q9n435 caenorhabdi
22	351	4.8	2768	5	Q9VC00 Q9vc00 drosophila
23	349	4.7	990	13	Q91803 Q91803 xenopus lae
24	344.5	4.7	3111	5	Q9VH10 Q9vH10 drosophila
25	341.5	4.6	1422	6	Q95KM4 Q95kM4 canis famli
26	340	4.6	2703	5	Q9VPE9 Q9vPE9 drosophila
27	340	4.6	5322	5	Q9N164 Q9n164 drosophila
28	339	4.6	2722	5	Q91935 Q91935 caenorhabdi
29	338.5	4.6	2157	11	Q9ZIR1 Q9zIR1 mus musculu
30	336.5	4.6	4498	13	Q93291 Q93291 fuqu rubrip
31	336	4.6	1062	5	Q960C4 Q960C4 drosophila
32	335.5	4.5	933	4	Q9UQ14 Q9uQ14 homo sapien
33	335.5	4.5	1026	4	Q9UJ57 Q9uJ57 homo sapien
34	333	4.5	1108	5	Q9ND10 Q9nd10 babesia big
35	333	4.5	1262	4	Q9UQ40 Q9uQ40 homo sapien
36	333	4.5	1343	4	Q9H7N4 Q9h7N4 homo sapien
37	332.5	4.5	1151	5	Q9VAV4 Q9vAV4 drosophila
38	331	4.5	2897	5	Q9VID9 Q9viD9 drosophila
39	331	4.5	2951	5	Q9W3Z0 Q9w3Z0 drosophila
40	330	4.5	1430	5	Q9W0B0 Q9w0B0 drosophila
41	329	4.5	1160	11	Q9ESC8 Q9eSC8 mus musculu
42	328	4.4	2701	4	Q9Y520 Q9y520 homo sapien
43	327	4.4	2187	11	P70670 P70670 mus musculu
44	326.5	4.4	1341	4	Q9UKV3 Q9ukV3 homo sapien
45	325.5	4.4	1178	5	Q9NHX5 Q9nHX5 drosophila

ALIGNMENTS

RESULT 1
P97868 PRELIMINARY; PRT; 1591 AA.
ID AC P97868; P70287;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE RETINOBLASTOMA BINDING PROTEIN 6 (PROLIFERATION POTENTIAL-RELATED PROTEIN) (P53-ASSOCIATED CELLULAR PROTEIN) (PACT).
GN RBBP6 OR P2P-R.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY (ISOFORM 2).
RC STRAIN-BALB/C; PubMed=9037032;
RX MEDLINE=97188447; PubMed=9037032;
RA Witte M.M., Scott R.E.;
RT "The proliferation potential protein-related (P2P-R) gene with domains encoding heterogeneous nuclear ribonucleoprotein association and Rb1 binding shows repressed expression during terminal differentiation.";
RT Proc. Natl. Acad. Sci. U.S.A. 94:1212-1217(1997).
RL [2]
RN REVISIONS.
RP STRAIN-BALB/C;
RC Witte M.M., Scott R.E.;
RX MEDLINE=97163396; PubMed=9010216;
RA Simons A., Melamed-Bessudo C., Wolkowicz R., Sperling J., Sperling R., Eisenbach L., Rotter V.;
RT "PACT: cloning and characterization of a cellular p53 binding protein that interacts with Rb.";
RT Oncogene 14:145-155(1997).
CC -!- FUNCTION: BINDS TO Rb1 THROUGH THE Rb1 POCKET DOMAIN. ASSOCIATES WITH HNRNP. ALSO BINDS TO P53. MAY BE INVOLVED IN GROWTH CONTROL

CC	AND DIFFERENTIATION.
CC	- I- SUBCELLULAR LOCATION: NUCLEAR.
CC	- I- ALTERNATIVE PRODUCTS: TWO ISOFORMS; ISOFORM 1 (SHOWN HERE) AND
CC	ISOFORM 2; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC	- I- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN TESTIS.
CC	MODERATE LEVELS FOUND IN HEART, LUNG, LIVER AND SKELETAL MUSCLE.
CC	VERY LOW LEVELS DETECTED IN KIDNEY, BRAIN AND SPLEEN. ALSO
CC	EXPRESSED IN GROWING UNDIFFERENTIATED MESENCHYMAL STEM CELLS BUT
CC	NOT DURING THE TERMINAL DIFFERENTIATION PHASE.
CC	- I- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR	EMBL; U83913; AAC72432.1; -.
DR	EMBL; U28789; AAB49620.1; -.
DR	MGD; MGI:894835; Rbtp6.
DR	InterPro; IPR001841; Znf_Cing.
DR	Pfam; PF00097; zf-C3HC4; 1.
DR	SMART; SM00184; RING; 1.
DR	PROSITE; PS00518; ZINC_FINGER_C3HC4; FALSE_NEG.
KW	Zinc-finger; Metal-binding; DNA-binding; Nuclear protein;
KW	Alternative splicing.
FT	ZN_FING
FT	61 101
FT	DOMAIN
FT	481 540
FT	1528 1585
FT	VARSPIC
FT	454 487
FT	CONFLICT
FT	55 55
FT	CONFLICT
FT	118 119
FT	CONFLICT
FT	142 142
FT	CONFLICT
FT	219 219
FT	CONFLICT
FT	222 222
FT	CONFLICT
FT	381 381
FT	CONFLICT
FT	396 396
FT	CONFLICT
FT	416 423
FT	CONFLICT
FT	430 430
FT	CONFLICT
FT	437 437
FT	CONFLICT
FT	448 448
FT	CONFLICT
FT	490 490
FT	CONFLICT
FT	504 504
FT	CONFLICT
FT	590 590
FT	CONFLICT
FT	742 742
FT	CONFLICT
FT	757 759
FT	CONFLICT
FT	764 764
FT	CONFLICT
FT	779 779
FT	CONFLICT
FT	783 783
FT	CONFLICT
FT	813 813
FT	CONFLICT
FT	1091 1091
FT	CONFLICT
FT	1117 1117
FT	CONFLICT
FT	1365 1365
FT	CONFLICT
FT	1382 1382
FT	CONFLICT
FT	1392 1392
FT	CONFLICT
FT	1397 1397
FT	CONFLICT
FT	1405 1405
FT	CONFLICT
FT	1591 AA; 17763 MW; 7D0C6798FA9D3748 CRC64;
QO	SEQUENCE

	Query Match	82.1%	Score 6054	DB 11	Length 1591
	Best Local Similarity	84.7%	Pred. No. 1.3e-309		
	Matches 1199	Conservative 31	Mismatches 111	Indels 74	Gaps 10
QY	1 MMEVDPNMKGAMLTNTCKYAIPTIDAEAYAIAGKREKPPFLPEEPPSSSEEDDPIPAELL 60				
DB	1 MMEVDPNMKGAMLTNTCKYAIPTIDAEAYAIAGKREKPPFLPEEPPSSSEEDDPPDELL 60				
QY	61 CLICKDITMDAVYIPCCGSSCDECIPTTLLESDKHCTPTCHQNDVSPDALANKFLRQA 120				
DB	61 CLICDITMDAVIPCCGSSCYDECIKRLALLESDEHTCTCHQNDVSPDALANKFLRQA 120				
QY	121 VNNFKENGYTRKRLKQLPFLFLVPPRPPLISQRLNLPSSRPILROODPVYERTVSP 180				
DB	121 VNNFKENGYTRKRLKQLPFLFLVPPRPPLISQRLNLPKMRSPISRQDDPLMIPTSSSA 180				
QY	181 CSDTKAGSCSDSGTSLRLPAPVISLTSNOSSLAPVVGNSAPAPVPDITATVISIV 240				
	181 HS-----APVISLTSNPALAPVSGNGSSAPAPVPDITATVISIS 222				

QY	241	HSEKSDGPEFRSDNKKLLPAAALTSSEHKSASSIAITAMLEKXG----	VPGTSPMNSIFVG	296
Db	223	HSEKSDGPEFRSDNKKLLPAAALTSSEHKSASSIAITAMLEEKXGYOVLGTPS---	LLG	278
QY	297	QSLHLGOLIPTTGVVRINARPGGRRPGWHSNKLGLYVSPQOIRNGERSCYRSINGR		356
Db	279	QSLHLGOLIPTTGVVRINARPGGRRPGWHSNKLGLYVSPQOIRNGERSCYRSINGR		338
QY	357	HHSERSQRTQSPLPAPPCVVPVPPPLYPPLPHTLPLRPVPPQFSPQFSSQPPAG		416
Db	339	HHSERSQRTQSPLPAPPCVVPVPPPLYPPLPHTLPLRPVPPQFSPQFSSQPPAG		398
QY	417	YSVPBGPPAPANISTACEPCVPYPAHSNTWPTQAPLISREEFYRQD-----		465
Db	399	YSVPBGPPAPANISTPWPVSSGVOIASHNTIPTTQAPLISREEFYRQDRRLKEEEKKS		458
QY	466	-----NDKGE-----SKFPYSSSSTRSSTYDS-SQGLAQHIALTL	501	
Db	459	KLDEFTNDFAELMEYKRIKERRRSTRSRKSPYSSSTRSSTYDSKRSQSTRSRYS	518	
QY	502	LSPSAHTLIDLHDPHPR-----PEEAERASAMIVHPDLMDIAHRSNRPARYRSR	557	
Db	519	RSFSRHSRSTRSNRPPLPRRGCRKSRRYNSRSH-----GYHRSRNSPPLRKHSHSR	573	
QY	558	SPPEFRGOSPTKRNVPPEEKEREYFNRYREVPVPIKAYGRSVDFRDPFEKERYRME	617	
Db	574	SPQAFRGOSPTKRNVPGETEREYFNRYREVPVPIKAYGRSVDFRDPFEKERYRME	633	
QY	618	RKYREMYEYKYGVAVAQRPSPANRDNDSPEELLPLNTRNSPFRGRREDYAAQSHRN	677	
Db	634	RKYREMYEYKYGVAVAQRPSPANRDNDSPEELLPLNTRNSPFRGRREDYAAQSHRN	693	
QY	678	RNLGNPPEKLTSDSINADNKRSEKESEENYPGCGKNKHKRRKRNEKEGESESE	737	
Db	694	RNLGNPPEKLTSDSINADNKRSEKESEENYPGCGKNKHKRRKRNEKEGESESE	750	
QY	738	LNPELLSTRKRCGSSGIDETKTDTLFLVLRSDRATPVDEPMDAESTFKVSPDKRE	797	
Db	751	LNPELLSTRKRCGSSGIDETKTDTLFLVLRSDRATPVDEPMDAESTFKVSPDKRE	810	
QY	798	KDRPKVSDKTKRRSDSAFAKKDNVLPKSGQOEYVDQDREKSRSEPLLAKAEEATK	857	
Db	811	KDRPKVSDKTKRRSDSAFAKKDNVLPKSGQOEYVDQDREKSRSEPLLAKAEEATK	870	
QY	858	IDSVKPSSSOKDEKVTGTPRAKHSASAKDTRQOSPTRBSKRYVPKTSOSQOPVTR	917	
Db	871	IDSVKPSSSOKDEKVTGTPRAKHSASAKDTRQOSPTRBSKRYVPKTSOSQOPVTR	930	
QY	918	RPRSLRKNVTLIA-EKNEBEKRRKSDVQDFEESMSKISKVGTELYVSPRKKKEGYE	976	
Db	931	KAKPEKKNKLDSGEXRRKKTTEKESYVDQDFEESMSKISKVGTELYVSPRKKKEGYE	990	
QY	977	KLERTPPEKDIASSTPAPKIKILNRETGKKIGNAENASTTKEPSELESTSIKQOEYK	1036	
Db	991	KLERTPPEKDIASSTPAPKIKILNRETGKKIGNAENASTTKEPSELESTSIKQOEYK	1050	
QY	1037	GKARKKVAAGSGSSSTLYDTTSTTSSTGSGVYRKSEKTTKTKTVIKTMEYVNDNTAAE	1096	
Db	1051	GKARKKVAAGSGSSSTLYDTTSTTSSTGSGVYRKSEKTTKTKTVIKTMEYVNDNTAAE	1110	
QY	1097	DVIIMIOVPOSKMDKDFESEEBEDVTTQPIQVSGPSSIIKNVTTKSATAKYTEKESE	1156	
Db	1111	DVIIMIOVPOSKMDKDFESEEBEDVTTQPIQVSGPSSIIKNVTTKSATAKYTEKESE	1170	
QY	1157	QPEKLOKILPKVEASHLMOHELARSSKSGASSEKGRADREHSGEKNDPKRRKGAQDKE	1216	
Db	1171	QPEKLOKILPKVEASHLMOHELARSSKSGASSEKGRADREHSGEKNDPKRRKGAQDKE	1230	
QY	1217	STVORLSGQHFCTLSSSKETRTSKHESVYRSGSKDTPGRDKKVVYDSDVSSKRR	1276	
Db	1231	STVORLSGQHFCTLSSSKETRTSKHESVYRSGSKDTPGRDKKVVYDSDVSSKRR	1290	
QY	1277	DERDELARRKQSPRGESLSGQKSKILREBRDLPPKKGASKSSNSGSPDDKKPHDHAKPY	1336	

5	1277	DERBELARRKSPRCGETESLGGOKSKLIREBRDLPKKGAESKSNSSNPDPDKKPHHKAHPY	1336
QY	1277	DERBELARRKSPRCGETESLGGOKSKLIREBRDLPKKGAESKSNSSNPDPDKKPHHKAHPY	1336
Db	1231	STVRLSEOGHFKLSSSSKERTSETSEHSEVROSSNKDTPDRDKVYDSDYSSSSARR	1290
QY	1217	STVRLSEOGHFKLSSSSKERTSETSEHSEVROSSNKDTPDRDKVYDSDYSSSSARR	1276
Db	1171	QPEKLOKPRKASHLMOHLBRSKSGSASSEKRAODRHSSEKDNPDKRSQAPKE	1230
QY	1157	QPEKLOKPRKASHLMOHLBRSKSGSASSEKRAODRHSSEKDNPDKRSQAPKE	1216
Db	1111	DVIIMIOVPOSKMKDPDEESEEDVTTOP IQSGVSPSSIIKNVTTKPSATAKYTEKESE	1170
QY	1097	DVIIMIOVPOSKMKDPDEESEEDVTTOP IQSGVSPSSIIKNVTTKPSATAKYTEKESE	1156
Db	1051	GKARKVAGSSGSSSTLVDTSTSSGSGPVKRSSEKTDTKRVIKIMEYNNNTAPAE	1110
QY	1037	GKARKVAGSSGSSSTLVDTSTSSGSGPVKRSSEKTDTKRVIKIMEYNNNTAPAE	1096
Db	991	KLEKTPERKDIASSTYPAKKIKINRELGKIGIAENASTTKPESEKLESTSSK IQOEKVK	1050

Db 1291 DERGELARRKSDPPRGKESLSQSGSKLREERDLPKKGAESKSNSSPPDKKHPHKAPY 1350
QY 1337 ETKRCEETKPPVDKNSGKEREKHAABARNGKSSG 1371
Db 1351 ETKRCEETKPPVDKNSGKEREKHAABARNGKSSG 1385
RESULT 2
ID Q96PH3 PRELIMINARY; PRT; 1616 AA.
AC Q96PH3;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROLIFERATION POTENTIAL-RELATED PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Camargo A.A., Moreira E.S., Simpson A.J.G.;
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF352051; AAL05625.1; -
SQ SEQUENCE 1616 AA; 182248 MW; D89F5BA4BED95D72B CRC64;

Query Match 73.9%; Score 5451.5; DB 4; Length 1616;
Best Local Similarity 75.3%; Pred. No. 5,6e-278;
Matches 1073; Conservative 92; Mismatches 167; Indels 93; Gaps 16;

QY 1 MNEVDPNNKGMINTGKYAIPITDAEYAIKGKKEPPFLDEBSSSEEDDPPEALL 60
Db 23 MNEVDPNNKGMINTGKYAIPITDAEYAIKGKKEPPFLDEBSSSEEDDPPEALL 82
QY 61 CLICDIMDAVYIPCCGSSCDECIRTLLESKDHTCTCHONDVSPALILANKFLROA 120
Db 83 CLICDIMDAVYIPCCGSSCDECIRTLLESKDHTCTCHONDVSPALILANKFLROA 142
QY 121 VNNFKNETGYTRLRKOLPFLFLVPPPLSQRLQPRSPILNQDPVYFRYTSPT 180
Db 143 VNNFKNETGYTRLRKOLPFLFLVPPPLSQRLQPRSPILNQDPVYFRYTSPT 202
QY 181 CSDTLAGSCSSGTLRLPAPSSISLTSSQSLAPPVGNSAPAPVPDITATYSIV 240
Db 203 -----HPAPSSISLTSSQSLAPPVGNSAPAPVPDITATYSIV 244
QY 241 HSEKSDGPPRSDDNKLLPAAALTSEHSKASSIATATLMEKG---VGTSPWNSIFVG 296
Db 245 HSEKSDGPPRSDDNKLLPAAALTSEHSKASSIATATLMEKG---VGTSPWNSIFVG 300
QY 297 QSLHGGQLIPTTGVPYRINARPGGRGWEHSNKLGLYLVSPQIIRGRSCYRSINRG 356
Db 301 QSLHGGQLIPTTGVPYRINARPGGRGWEHSNKLGLYLVSPQIIRGRSCYRSINRG 360
QY 357 HHSERSQRTQSLSLATPCGFVYPPPLPLPPPHILPLPGVPPPPPSQFPSSQPTAG 416
Db 361 HHSERSQRTQSLSLATPCGFVYPPPLPLPPPHILPLPGVPPPPPSQFPSSQPTAG 420
QY 417 YSVPPGPPFPAANISTACFSPGVPFAHSNTMPTQAPLSREEFREBO-----465
Db 421 YSVPPGPPFPAANISTACFSPGVPFAHSNTMPTQAPLSREEFREBO-----480
QY 466 -----NDKGR-----SKPPYSGSSYSSRSTYDS--SQGLAQHIALTL 501
Db 481 KLDEFNDFAKELMEYKTIQKERRRSFSRSKSPYSSGSSYSSRSYTSKRSRSTRSRSY 540
QY 502 LSPSAHTITDLHDHP-----PEEAARSAMVHMPDLMDIAHAASRSPYRYRSR 557
Db 541 RSFSSRSHSSYSRSPYPRGRGKSNYRSRSH-----GYHRSRSHSPYRYRSR 595
QY 558 SPPEFGSGPTKRNVPREKEREYFNRYREVPPPYDIKAYYGSVDFRDPFEKERYREWE 617

Db 596 SPQAFRGSGPNKRNVPQGETEREYFNRYREVPPPYDMKAYYGSVDFRDPFEKERYREWE 655
QY 618 KRYREMYEYTYGYAVGAQPPRSANREDFSPERLLPLINRSPFTRGREDYAAAGOSHAN 677
Db 656 KRYREMYEYTYGYAVGAQPPRSANREDFSPERLLPLINRSPFTRGREDYAAAGOSHAN 715
QY 678 RNLGNYPEKLTSPRSHNANKDNPKESESENVPGDGKGNKKHKKRRNEKEGESESF 737
Db 716 RNLGNYPEKLTSPRSHNANKDNPKESESENVPGDGKGNKKHKKRRNEKEGESESF 772
QY 738 LNPelletSRKCRGSGIDETKTDLTVLPSSRDAATPVDEPDMAESTIFKSVSDKRE 797
Db 773 LNPelletSRKCRGSGIDETKTDLTVLPSSRDAATPVDEPDMAESTIFKSVSDKRE 832
QY 798 KRPKYKSKTRKSDGSAATKADNVLAASKGPOEKVDGDRKSPRSEPLKAKAEATK 857
Db 833 RDKPRAKGGKTRKKNKSVSKKENYVPAKGPQOEVDGERRSRSPSEPIKAKEETPK 892
QY 858 IDSVPRSSSOKDEKVTGPRKAHSAKADTRRQSPRTRSKRTVPKTSQKSOQVPR 917
Db 893 IDSVPRSSSOKDEKVTGPRKAHSAKADTRRQSPRTRSKRTVPKTSQKSOQVPR 950
QY 918 RPSLTKIYLLIAREKN-----EREKRR---KSYDKDFESSSMKISVEGETEIVKPS 967
Db 951 EKA-KRPY-----EKNKFLDNKGGKRRKTEKGVKDFESSSMKISLVEYTEIVKPS 1004
QY 968 KRMGDEVKLETPKDKIAGSTTPAKIKILNREGKIKGNAENASTKEPSEKLEST 1027
Db 1005 KRMGDEVKLETPKDKIAGSTTPAKIKILNREGKIKGNAENASTKEPSEKLEST 1063
QY 1028 SKIKOEKVGKAKRKVAAGSSSTLVDTSTSGSGVPRKSEETDTRVYKMEY 1087
Db 1064 SKIKOEKVGKAKRKVAAGSSSTLVDTSTSGSGVPRKSEETDTRVYKMEY 1123
QY 1088 NNDNTAPADVYIIMIQVPSKMDKDFESEEDVKTQPIQSGKSSIIKNTYTPSAT 1147
Db 1124 NNDNTAPADVYIIMIQVPSKMDKDFESEEDVKTQPIQSGKSSIIKNTYTPSAT 1183
QY 1148 AKYTEKESQPEKLOKLPREASHLMOHELRSKSGASSSEKGRADREHSGSEKNDPKR 1207
Db 1184 AKYTEKESQPEKLOKLPREASHLMOHELRSKSGASSSEKGRADREHSGSEKNDPKR 1243
QY 1208 KSGAQDKESTYDRLSEOGHFKTLSSSKETRSSEKHESEVRSNNDPFGDKYDVDS 1267
Db 1244 KSGAQDKESTYDRLSEOGHFKTLSSSKETRSSEKHESEVRSNNDPFGDKYDVDS 1303
QY 1268 RYSSSKRRDERGELARRKSDPPRGKESLSQSGSKLREERDLPKKG-AESKKSNSPPRD 1326
Db 1304 RYSSSKRRDERGELARRKSDPPRGKESLSQSGSKLREERDLPKKG-AESKKSNSPPRD 1363
QY 1327 KRPDHKADYETKRCCEETKPPVDKNSGKEREKHAABARNGKSSG 1371
Db 1364 KRPDHKADYETKRCCEETKPPVDKNSGKEREKHAABARNGKSSG 1408
RESULT 3
ID Q15290 PRELIMINARY; PRT; 948 AA.
AC Q15290;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RB PROTEIN BINDING PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96129310; PubMed=8595913;
Sakai Y., Saijo M., Coelho K., Kishino T., Nishikawa N., Taya Y.;

"cDNA sequence and chromosomal localization of a novel human protein, RT RBQ-1 (RBQB6), that binds to the retinoblastoma gene product.";
RT Genomics 30:98-101(1995).
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR EMBL: X85133; CA59445.1;
DR InterPro: IPR001841; Znf_ring.
DR Pfam: PF00097; zfc-C3HC4.1.
DR SMART: SM00184; RING; 1.
KW Zinc-finger.
SQ SEQUENCE 948 AA; 107156 MW; 5F24BDE5E9340A2 CRC64;

Query Match 47.5%; Score 3500.5; DB 4; Length 948;
Best Local Similarity 73.2%; Pred. No. 7.2e-176;
Matches 691; Conservative 44; Mismatches 130; Indels 79; Gaps 10;

1 MMEVKDPNNKGMALITNGKAAIPTIDAEEYAIGKKEKPPFLPEPSSSSSEDDPIAEILL 60
23 MMEVKDPNNKGMALITNGKAAIPTIDAEEYAIGKKEKPPFLPEPSSSSSEDDPIAEILL 82
61 CLICKIMTDVAVYIPCCGSSCDECIPTILLESDEKTCPTCHQNDVSPALLANKFLRQA 120
83 CLICKIMTDVAVYIPCCGSSCDECIPTILLESDEHTCPTCHQNDVSPALLANKFLRQA 142
121 VNNFKNETGYTKRLKQOLPEFLVPPRPLSQNLQPSRSRSPILRQDPVFRVTVSPT 180
143 VNNFKNETGYTKRLKQOLPEFLVPPRPLSQNLQPSRSRSPILRQDPVFRVTVSPT 202
181 CSPTKTATGSCSDSGLSRLLPAPSISSLTSSNOSSILAPVSGNSPAPVPDITATVSI 240
203 -----HPAPSISSLTSSNOSSILAPVSGNSPAPVPDITATVSI 244
241 HSEKSGPPRDSNKLPAALTSSEKSGASTAITALMEKG-----VPSTSWNSIFVG 296
245 HSEKSGPPRDSNKLPAALTSSEKSGASTAITALMEKGYQVPLTSPS---LLG 300
297 QSLHGOILPTTGPVRINARPGGPGMGWESHKGLGYLSPPOIRGERSCYRSINRGR 356
301 QSLHGOILPTTGPVRINARPGGPGMGWESHKGLGYLSPPOIRGERSCYRSINRGR 360
357 HHSESRQRTQSPSLPATPCFVYVPPPLYPPTHTLPLPGVPPPOFSPQFPSSQPTAG 416
361 HHSESRQRTQSPSLPATPCFVYVPPPLYPPTHTLPLPGVPPPOFSPQFPSSQPTAG 420
417 YVPPPPPPAPANISACSPGVPTAHSMPTTAPLISREEFYEQ-----465
421 YVPPPPPPAPANISACSPGVPTAHSMPTTAPLISREEFYEQRLKEEEKKS 480
466 -----NDKGRE-----SKFYSGSSYSRSTYDS-SQGLAQHIALT 501
481 KLDEFTNDPAKELMEYKIKIKERRRSRSKSPYSGSSYSRSTYDSKSGSTRSRYS 540
502 LSPSAHTLDDLHDHPH---PPEAARSAMLYHMPDLMDIAHARRSPRYRYSR 557
541 RPSRSRSHRSYSRSPYPRRGKRSRNYRYSRSH---GYHRSRSPRYRYSR 595
558 SPPEFRQSTPKRVNPREEREYEFNRYRVPYDIAKAYGSRVDRDPEKRYEME 617
596 SPFAFRQSTPKRVNPREEREYEFNRYRVPYDIAKAYGSRVDRDPEKRYEME 655
618 RYKREWEKYYKGYAAGORPANSANREDSPERLLPLINRSPTRGRRREDYAGAGSHN 677
656 RYKREWEKYYKGYAAGORPANSANREDSPERLLPLINRSPTRGRRREDYAGAGSHN 715
678 RNLGANTPEKLRSDSHNAKDNPKSEKESBENYPGDGKGNKHKRRKRNKEKGESESF 737
716 RNLGANTPEKLRSDSHNAKDNPKSEKESBENYPGDGKGNKHKRRKRNKEKGESESF 772
738 LNPBLETSRKRGSSGIDETKTDTLVLSRDDATPVYRDEPMDAESITTKSVSDKCKRE 797
773 LNPBLETSRKRGSSGIDETKTDTLVLSRDDATPVYRDEPMDAESITTKSVSDKCKRE 832
798 KDKKVVSDKTKRKSDDSAATAKKDNVLPKSGQEKVDGDEKPSRBPPLKAKKEATK 857

833 RDKPRAKGDKTRKNDGSAVSKENIVKPAKGPQEKVGDVBDLIDLNLQKKPRRLR 892
858 IDSVAPSSSSQODEVYTGTPRRAHGSKAKDTORQSPRTFRSKR 901
893 LVLNHHLLPLRRKKSLEPEKL-----TLNQQKTPRKNKTSOR 930

RESULT 4
ID 09H5M5 PRELIMINARY; PRT; 628 AA.
AC 09H5M5;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)
DE CDNA: FLJ23301 FIS, CLONE HEP1120.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hiki J. T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RA "NEO human cDNA sequencing project."
RT Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AK026954; BAB15600.1;
SQ SEQUENCE 628 AA; 70907 MW; F9FD5998518CC9A CRC64;

Query Match 23.1%; Score 1702; DB 4; Length 628;
Best Local Similarity 78.6%; Pred. No. 1.1e-81;
Matches 331; Conservative 40; Mismatches 48; Indels 2; Gaps 2;

952 MKTSYEGEPIVAPSRKRMGEVLEKLEPTEKDKTIASTTPAKTKIKLNREGTKIGNAE 1011
1 MKTSYEGEPIVAPSRKRMGEVLEKLEPTEKDKTIASTTPAKTKIKLNREGTKIGNAE 1011
1012 NASTKEPESEKLESTSSKIKOEKVKAKRKAAGSESSSTLVDTYSTSGSPYRKSE 1071
60 NISNTEPESEKLESTSSKIKOEKVKAKRKAAGSESSSTLVDTYSTSGSPYRKSE 119
1072 EKTDTKRTVYKIMEYVNDTAPAEVYIMIQVPSKMDKDFESEEDVYKTOPQSVG 1131
120 EKTDTKRTVYKIMEYVNDTAPAEVYIMIQVPSKMDKDFESEEDVYKTOPQSVG 119
1132 KPSSITIKNTTKSATPAKTEKESQEPKLOXLPKASHHELMQHELRSKGSASSEKGRA 1191
180 KPASVILKINVSSTKPSNIVKYPEKESQEPKLOXLPKASHHELMQHELRSKGSASSEKGRA 1239
1192 KOREHSGSEKDNPDKRSKGAQPDKESTVDRLSQEGHEKTLSSQSKETRTSEKHESVRGS 1251
240 KDRDYSLKEENPEKRNSTQPEKESLMDLNLQGNKSLSSQSKARTSDKHSTRASS 299
1252 NKDFTPGRDKVYDSDHYSKRRDERGELARRKDSPPRGKSLSSQSKLLEERLPK 1311
300 NKDFTPGRDKVYDSDHYSKRRDERGELARRKDSPPRGKSLSSQSKLLEERLPK 1359
1312 KG-AESKSSNSPPROKPPHDKAPYITKTRCPERTKTVYDNGSGERKHAABEANGESS 1370
360 KGCGDSKSSNSPPROKPPHDKAPYITKTRCPERTKTVYDNGSGERKHAABEANGESS 1419
1371 G 1371
420 G 420

RESULT 5
ID 09NPX4 PRELIMINARY; PRT; 529 AA.
AC 09NPX4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)


```

DR 01-OCT-2000 (TREMBlrel. 15, last sequence update)
DE 01-OCT-2000 (TREMBlrel. 15, last annotation update)
DE HYPOTHEICAL. 59.9 KDA PROTEIN (FRAGMENT).
GN DKFPP761B2423.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=AMYGDALA;
RA Blocker H., Boecher M., Brandt P., Mewes H.W., Well B., Wiemann S.;
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AL359564; CAB94869.1; -.
KM Hypothetical protein.
FT NON_TER
SQ SEQUENCE 529 AA; 59897 MW; 8F5A66CDF74CCA66 CRC64;

Query Match 17.7%; Score 1305.5; DB 4; Length 529;
Best Local Similarity 77.9%; Pred. No. 5,5e-61;
Matches 250; Conservative 32; Mismatches 38; Indels 1; Gaps 1;

QY 1052 TIVDTYSTGTSGSPYRKSEKEDTKRTYIKTMEEYNNNDTAPADYIIMIOVQSKWMD 1111
DB 1 TIVDTYSTGTSGSPYRKSEKEDTKRTYIKTMEEYNNNDTAPADYIIMIOVQSKWMD 60
QY 1112 DFESEEDVKTQPIQVSGKPSIIKNTYTPRSATAKYTEKSEDPKILQKLPKASHE 1171
DB 61 DFESEEDVKTQPIQVSGKPSIIKNTYTPRSATAKYTEKSEDPKILQKLPKASHE 120
QY 1172 LM0HELRSSKSGASSKSGAKRHSSEKNDKRSAGQPKESTVRLSLQSGHFKTL 1231
DB 121 LI0HELRSSKSGASSKSGAKRHSSEKNDKRSAGQPKESTVRLSLQSGHFKTL 180
QY 1232 SOSSEKTRSEKHEVYSGSSNKDFTPGRDKYVDSDYSSSKRRDREGELARRKDSPPR 1291
DB 181 SOSSEKTRSEKHEVYSGSSNKDFTPGRDKYVDSDYSSSKRRDREGELARRKDSPPR 240
QY 1292 GRESLSGOKSKLREERDLPRKG-AESKSSNSPPRDKPHDRKAPYETKRCPEETRPVDK 1350
DB 241 NDSASGOKSKLREERDLPRKG-AESKSSNSPPRDKPHDRKAPYETKRCPEETRPVDK 300
QY 1351 NSGKERKHAAREARKNESSG 1371
DB 301 NSGKERKHAAREARKNESSG 321

RESULT 6
QY 09XZ21 PRELIMINARY: PRT: 1231 AA.
AC 09XZ21;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE BCDA:ID21643 PROTEIN.
GN BCDA:ID21643 OR CG3231.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Gelunker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,

```

```

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacble J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster." ;
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,
RA Agbayani A., Arcaina T.T., Baxter E., Blazey R.G., Butenhoff C.,
RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,
RA Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,
RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Pacble J.M.,
RA Park S., Sequeira A., Sethi H., Snir E., Svitskas R.R., Weinburg T.,
RA Carliker S.E.;
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR EMBL: AE003463; AAF47162.1; -.
DR EMBL: AF132177; AAD34765.1; -.
DR FLYbase: FBgn027522; BCDA:ID21643.
DR InterPro: IPR001878; znf_CCHC.
DR InterPro: IPR001841; znf_ring.
DR Pfam: PF00097; zf-C3HC4; 1.
DR Pfam: PF00098; zf-C3HC4; 1.
DR SMART: SM00184; RING; 1.
DR SMART: SM00343; znf_C2HC; 1.
DR SMART: SM00343; znf_C2HC; 1.
KW Hypothetical protein; Zinc-finger.
SQ SEQUENCE 1231 AA; 139038 MW; 4A91FF172BDEAB0 CRC64;

Query Match 7.4%; Score 542; DB 5; Length 1231;
Best Local Similarity 21.4%; Pred. No. 1.7e-20;
Matches 301; Conservative 188; Mismatches 434; Indels 486; Gaps 63;

QY 35 KEKPEFLPER-----SSSEEDDPIPAELLCLICKIDMTDVAIVIPCCGSSCDCICRT 88
DB 188 ROKPDAENNSADPVLPAVONOE---IPEDLICIGRIDIFVDAVMIPCCSSSCDCDCVRT 244
QY 89 TLLESKHTCPTCHQNDVSPDALIANKFLRQAVNNFNKNETGYTKRLKQLPPLFLVPPP 148
DB 245 SLIESESECPDCKEKKNKSGSLIPNRLFNVSNAFKNKNGYKNSAK----- 292
QY 149 RPLSQNLQPKRSPLIR--QDDPVVFRYVSPGCDTKAGSCDSGSLTRPAPSIS 206
DB 293 -PAAVRN--EERPVKEVEKRPVA---EVEDETEVK-----PEKOKES 331
QY 207 ITSNOSLIAPVGVNSAPAPVDPITATVSVHSEKSDGPRDSDNKLLPAAALTSEH 266

```

```
D 332 ETN-----GNNPKSESEPPPTATPS-QKEKKYDSDYEDNITIKMPQPAADS-- 379
Q 267 SKGASAIATALMEKGVPTSPMNSIFVGOSLHGOLIPTTGVRINAAARPGGPGME 326
D 380 -----TTVPKRS-----PQYS 391
Q 327 HSNKGLVSPPOQIRREBRCYSINGRHSEKSOQTOSPILATCFVVPVPPPLP 386
D 392 HRSSESH-----RRDRSDYVDHDKH--QRPKSESVNMDRS----- 427
Q 387 PPPHTLPPGVPPOQSPQSPSSOPPTAGXSVPPGPPAPANISTACFSPGVTASHN 446
D 428 -----LTPPIG----- 434
Q 447 TMTPTQABLSEEFYRONDNGREKPYGSSYSRSSYTDSSQGLAQHIALTLSPSA 506
D 435 TLPSYOGHMAESEEAR-----SSAYKPPYQMOQG-----PPP 469
Q 507 AHTLDLHDHHPPEAARSAIVHMDLMDIAHAKSRPP--YRKRYSRSPPEPRG 564
D 470 MH-----MMSHMPAYNNGFNMGQRPPLSYVYQONOSVHP----- 505
Q 565 QSPTRKRVPREKEREYENRYREVPPYDIKAYYGRSVDPDPPEKERYREMERKRYREMY 624
D 506 -----MRAPIG--SAGGGMNMMSQPQSPN-----L 530
Q 625 EKYKGYA--VGA-----OPRPSANREDSPERLLPLNIRNSPTTGREDEYAAGOSHNR 678
D 531 ASIYGAAKVGSGPIDPPLFAFNIMKEKRRKVKDRFRSDRHHSRDP--QGRHNFK 587
Q 679 NLGNYR--EKLSTRDSINAKD--NPKSEKESENVPGCGKNGKKH--RKRRENEKGEES 734
D 588 S-----PAYEKDNSD--NLMDKRRSRERKREH-----SYERNHRHRRSSROPDGSGS 635
Q 735 ESNLPELLETSRKCRGSGIDETKTDTLFLPSRDATVPREDPMOAEESTFKSVSDKD 794
D 636 PG-----GRIKR--SG-----HRSASP----- 651
Q 795 KREKDKPKVSKDKTKRKSDGSATAKDNVLPKSKG--POEKVVDGDRKSPRESEPLKRAK 852
D 652 -----KPGYKSDYDKPYNKPSAPKTEAVEPPPGFELQDLD--EDGYKNKHPTSEA 703
Q 853 EEARLIDSVKSSSSQKDEKVTGPRKAH-----SKAKDTRROSQPTTRSKRTVPR--T 906
D 704 SOSKSGSSKRGENRHE-----APRKRHSRSISKEKPPMDSNVRSLTTPAKITTPMT 759
Q 907 SSOSQOPVTRRPRSLRI--NYLLAREK-----NEREKKKSVXKDESSSMKISK 956
D 760 AAOURLQ--RESSPTPEKSHDDYLAKAKIMASOPVINDTE--METNNGEKKNKAS--PLSK 815
Q 957 VEGTEIVKPSPRKMEGVKLETPPEKDKTASSTTPAKITLNBETGKIKGNAENASTT 1016
D 816 -----DRKKKKDKKKAER-----KKNKKKRAKKEKGDQKSSSS 851
Q 1017 KEPEKELESTSSKIKQEKVKAKAKRYAGSGSSSTLYDYTSTSGSGPYRKSEKEDT 1076
D 852 VNRSDD-----SDINNSSLIMESNYKVLSPRAQSPSI-----EINAAQSLPHNATENNVP 901
Q 1077 KRYITKMEEYNNNTAPAEVYIIMIQVPOSKMDKDESESEEDVKTQPTQSGKPPSI 1136
D 902 KSHSILTVGAASDNLGRSKLSBANSVNLKWEIDENTILEG--SSKKAAGADDPSEI 960
Q 1137 IKNTTTPSATATKTEKESEQPEKLOKLPRKASHELMOHELRRSSKGSASSEKRAKDEH 1196
D 961 TSDVLRKAENMIFAKAINAIRMEFOYI-----INKKMSKDSRV 1000
Q 1197 SGSEKD--NPKRKSQAQPKESTVDRL-----SEQHEFTLSQSSKETRTSKHESVR 1248
D 1001 VRSKDKSSSSPRRNNSS-----SRSVKDLRTGKIKNSDRSRSDKSGRRRAARSDDDDANR 1055
Q 1249 GSSKDKPTGDKKLVYDSDRYS--SKRDERGELARAKDOSPGRKESLSQCKKLPRER 1307
D 1056 GRSDHSGSRKDNK-----SRDRAAPSEKROER--STKRSSP-----EDDKLRKON 1099
```

```
Q 1308 DLPRKGAESK-----KSNSSPPDDKKPHDHKAPYETK-----RPECETKP----- 1347
D 1100 ---KEQSESKHGKHQDQNNSSDSDRRAAKNTKSSDSRVSVYAVAPAPKPCRPDPNPFKE 1156
Q 1348 VDKNSGKER-----EKHAAEARNGKE 1368
D 1157 VDTSSSSSLVVKYDNTIQKEGASSDNGME 1185

RESULT 7
062235
ID 062235 PRELIMINARY; PRT: 1190 AA.
AC 062235;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE F36F2.3 PROTEIN.
CN F36F2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Cottege A.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; Pubmed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z81532; CAB04326.2; -.
SQ SEQUENCE 1190 AA; 137239 MW; B219D269962432AB CRC64;

Query Match 6.5%; Score 480.5; DB 5; Length 1190;
Best Local Similarity 19.0%; Pred. No. 2.8e-17;
Matches 276; Conservative 168; Mismatches 400; Indels 605; Gaps 51;

Q 1 MME--VKDPNKGMLNTGKAIPTI--DAEAYAIKKEKPPPLPEEPSSSEDDPIR 56
D 252 LMETTVDDPD---AMHPSGKYVIPIMHMKARQETLARKNE-----DGSSSPAOTSRRVP 303
Q 57 AELICLICQDINTAVYIPCCGSSCDECITITLESQKHCP--TCHQNVSDALIAN 114
D 304 PELLCPICQSLFEKAIYVSCCGNSYCADCEARILDPNOKCPGADCGK-DISTSTIIPN 362
Q 115 KFLQAVNNFNENGTGYKRLRKQPLPEFLVPPPRPLSQRMIAQRRSRPILROODPVVFR 174
D 363 KTLDAAMAAWLSATGCPGAPPTPOLV-----EPQQR 394
Q 175 YTVSPGSDITAGSCSDSGTSLRLPAPSSISLTSNOSSLAPPVSGNBSAPAPVPDITA 234
D 395 IRIQ-----LKAAPS-----SSOSQITP----- 411
Q 235 TVTISVHSEKSDGPFRRSDNKLPLPAALITSEHSKASAIATALMEKGVPTSPMNSIF 294
D 412 -----SGISGSLTV 421
Q 295 VGOSLHGOLIPTTGVRINAAARPGGPGMEHSNKLGLVYSPPOQIRRGERSYRSINR 354
D 422 QQQFTL-----TSVSS 432
Q 355 GRHHSERSQRTQSPSLPATCFVVP--PPPLYPPTHLLPLPPVPP--PQSPQPPSS 410
D 433 GTISIAQPSNVN--NPSIGIPLASQVPSMVODVSLPPPOLRQELRPGIGLPGF----- 484
Q 411 QPPTAGSVPPGPP-----PAPANISTACSPGVPYTAHSTMTPTQAPLPSRE 459
D 485 -----GLPPGVPGLSATVLPQHOHSMPLNYGMPLEFSAGPPRAVS-----SVPRPSAID 533
```



```
Db 1014 EXEKKSKTEBEAKKEKKKSDKKREKSE-----KRSKEKE-- 1053
OY 1115 ESEEDVKTTOPIQSGKSSIIKNVTKPSATAKYTEKESQPEKLOKLPKEASHE--- 1171
Db 1054 --ESRDLAKK-----KEEETK-----EKKESENHKSKEKEDKHEEDK 1091
OY 1172 -LMOHELRSKSGSASSEKGRANDREHSGSEK---DNPDKRKSGAOPDKESTYDRLSEQGH 1227
Db 1092 SKKKEEDKKEKKHHEESKRKKEEDKDKMEKLEDOKSNKKKEDKNEKKSQ-----H 1143
OY 1228 FTTLSOSKETRTSEKHEVSRRSSKNDFTPGRDKKYDYOSROYSSSKRDEGELARRKD 1287
Db 1144 VKLV--KRESDKKEKKEKNEKSEKTEKLEISSKQKNEVKKKESKSDOKKE----- 1194
OY 1288 SPPRGKESLGGOKSKLREBRDLPPKKGASEKSSNSPPRDKKPHDHAKAPYETKRCPEETRP 1347
Db 1195 -----KEMKESEKKL-----KKNEDRKOTSYEENKKOKETKK--EKNKPKDOKKN 1240
OY 1348 VDKNSGKER---EKHAAEARNGKES 1369
Db 1241 TTKSGGKKKESMESEKKAENOOKS 1265
```

```
RESULT 9
OY0Q35 PRELIMINARY: PRT: 2752 AA.
AC OY0Q35:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RNA BINDING PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtaki S., Umeki K., Sawada Y.;
RT "Homo sapiens mRNA for RNA binding protein, complete cds.";
RL Submitted (Jul-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016092; BAA83718.1; -
DR InterPro; IPR002965; P_tich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 2752 AA; 299672 MW; 109C64F181097123 CRC64;
```

```
Query Match 5.4%; Score 396.5; DB 4; Length 2752;
Best Local Similarity 20.2%; Pired. No.1.9e-12; Indels 577; Gaps 75;
Matches 362; Conservative 109; Mismatches 661;
```

```
OY 31 AIGKKEKPPLEPESSSEEDDPIPAELLCILCKIDMTDAVVI PCCGNSGCD----- 84
Db 48 ALVKRRPDLIDHEKRR-----VELRCLLEEMEE-----QGYEQOIOEK 90
OY 85 --CIRTTLESKDHCPTCHONDVSPDALIANKFLROAV-----NNFKMETGYTKR 133
Db 91 VAFPRILMLE-----KDVNPGKEETPGQRAVAVETHQLAELANKNE-----R 134
OY 134 LR-----KQLPPLFLVPPRPPPLGQRNQLQPPRSRSLNRQD 169
Db 135 LRFAFGISDYDGSFDPQRARAKQAP-----EPPKPYSLVRESSSSRSTPKOK 189
OY 170 PVFERYTVSPGSDTKTAGSCSDGT-----LSRLPAPSISS 206
Db 190 K-----KKKKRGKRSESSPRRERKSSKKKKHSESESKKKRHRSPTRK--- 235
OY 207 LTSNQSGLAPVSGNPSSAPAPVDITATVVISYHSEKSDGPFDPDNKLLPAAALTSEH 266
Db 236 --SKRKSCKDKKRKRSRSTPAD-----KSRRAHRSTSDASASSD-----TSR 276
OY 267 SKGASSTAIATAMEKKEVPGTSPWNSTFVGOSLLHGQLITPGVYRINAARPGGGRPGWE 326
```

```
Db 277 SRSRSAAKT---HTTALAGRSP-----SPASG-----RGEGDAPSE 312
OY 327 HSNKGLIYVSPPOQIRRGGRSCYRSTINRGHHSENQRTQSSLSLATPCFVVPVPPPL-- 384
Db 313 PGTSTQRPSSSEPTATKOPSSPYEDKDKKESATRPSPERSSTGCPAPAPPLAE 372
OY 385 ----YPPPHPLPL--PGVPPPOFS-----PQFPSSQPTAGSYVPPGPPAPANI 431
Db 373 RHGGSFQPLATTPPLSGEPVNPSPSEASPTDRSPKPSPEKLPSSSESSPPS--PQPTKV 430
OY 432 ST-ACFSPGVP-----TAHSNT----- 447
Db 431 SHNASSPESPAPAPAGSHREISSPTSKNRSHGAKRDKSHHTPSRRMGRSRSPATA 490
OY 448 -----MPTQAPLISSEETRYEOND-----KGR----- 470
Db 491 KRGSRSTRPTTGRGHSRSKSPQMRGRSAORMGSRSPQRGRSPRQPGMSRSNTOR 550
OY 471 --ESKFPYGSYSRSRSDYDSSQGLAONHTLSPSAHTLDLHDHPPEAEARSA 528
Db 551 RGRSRARRGRSHSRSPAT--RGRSRSTPARGRGRSRSTPARRRSRSTPTRRRSR 607
OY 529 MIVHMPDLMDIAHARSPPYR-----RYRSRSRSP-----PEFGOS--- 566
Db 608 TPARR-----GRSRSTRPARRSRTSPVRRRSRSPARRSGRSPRTPARGRSR 661
OY 567 -PYKRVPEEKEREYFNRYREVPYDIKAYGRSVDRDPE-----KEYY 613
Db 662 TPARRGRSRSTPARRSGRSRSTP-----ARRGRSR--RTPRGRSRSLVLRGRSHS 715
OY 614 REMERKRYREYKYYGYAVGAQPPSANREDSPERLLPLNTRNSPFTTGRREDYAAQ 673
Db 716 RFPQRRGSGSSSEKKNKSKTSQR--SKNSSP-----MKKRISRRRSLSR 766
OY 674 SHR-----NRNLGNYV--EKLST--RDSHNAKNPKSKE-----KES 707
Db 767 SKAKSLSLRSLSGSSPCQKQKQTPRRRSRSGSSQPKAKSRTPRRSRSSSSPPPKQ 826
OY 708 ENVPGCGKNNKHKKKRRREKEGESESLNPELLETS-----RCRSGSIDENKTDI 762
Db 827 SKTPRSQSHSSSPHPKVKSGTPPRQG--STTSQANEQSVTPQRRCFESPPPELKSRT 885
OY 763 LFLVPSR-----DDATPVDEPMDAES-----TFKSYSDKDKREKDKPKV 803
Db 886 ----PSRHSCSGSPPRVKSSTPRPOSPSRSSSPQKVKALIIIPRQRSHGSSSPSPSV 941
OY 804 KSDTKRKSS-----DGSATARKDNVLPKSGPQEKVDGDRKESPR--- 843
Db 942 TSTRTPRRSRSVSPCSNVEESRLPRYSHSGSSPDTRKVPKETPPROSHSISIPYKVK 1001
OY 844 -----SEPPKKAKKEATKIDSVKS-----SSSQKDK---VTGTPRKAS 882
Db 1002 QTPPGSLSGSKSPQEKSKDLSVQSCPSLSLCAGVKSTPGEISYFCVSSLOLKGOS 1061
OY 883 KSAKQTRPRQ--SOPRTRSKRTVPKTSQKQAPVTRPRSLRKINYLIAREKNEREK 941
Db 1062 QTSPPDRSDTSPFVQSHSESPLOSQKQTSPPKGRSRSSSVTELASPSPIRQDRGEF 1121
OY 942 SV-----DKDESSSMKISKVEGTEIVKPSPKRMEDVEKLETPPKDIASS 990
Db 1122 SASPMLKSGMSPQSRROQSSSYPTVDSNLSLQGSRLFEAESEK-EKMALPPOED-ATA 1178
OY 991 TTPAKKIKLN-----RETKKIGNAEANASTTKPESEKLESTSSKIKQ 1032
Db 1179 SPPROKDKFSPFVQDRPESSLVFKOTLRTPPRERSGAGSSPEKQNSALPTSSODEEL 1238
OY 1033 EYKYGKAKRKVAGSESSSTLVQYTTSTSGSPVRKSEKTDTKRTVITKMEYNNDMT 1092
Db 1239 MEYVESSEFAGOLISLSELEKMTSINFESS-----EVERPVSILIDQSGOAS 1292
OY 1093 APAEDVITIMIQVQ--SKWDKDFESEEDVKTTOPIQ--SVGKPSIIKNNVTKPSATA 1148
Db 1293 LEA-----VEVPSMASSWGCPHFSPEHKEL--SNSPLRENSFGSP--LEFRNSGPIGTE 1342
```

QY 1149 KYTEKESQPEK-----LQKLPKEASHLQHELRSSKSSASSEK----- 1189
DB 1343 MNTGSEVSEKEDLNGPFLQLETDLPDLKKEQSTRSS-GHSSSELSPDAVEKAGMSSNOS 1401
QY 1190 -----RAKDEHSGSE-----KD-----NPKR-KSGAOP---DKESTVDRLESG 1226
DB 1402 ISSPVLADVPRTPSRERSASSASPEMKODLPTPSRERSSSGSPGLRDSGTPSRISLSG 1461
QY 1227 H-----FKLISQ-----SSKETRTSEKHESVSGSS-----NKDETPOGRDKVDYD 1266
DB 1462 SSPGKMIDIPRTPEFRGSECDSSPEPKALPQTPRPSRSPSSPELNNKCLTPQREKSGSES 1521
QY 1267 SRDY-----SSSKRRD-----ERGL-----ARKKDPPEKESLSG- 1298
DB 1522 SYDQKTVATPVLQGRSSGSSQELDKPPASPOERSESSSPDKAKTPTPLRQSRSSGS 1581
QY 1299 -----QSKLR-----EERDLPKKGAKSKNSSPPDKKPHDKAPYETRRPCE 1343
DB 1582 SPEVSKSLSPRSRSGSSPEVKOKPRAAPRAQSGSDSPKAPAPRALPRRSRSGSS 1641
QY 1344 E-----TKPYDKNSGKERKHAALNKGKSSGANCHVYLTROTLPWRRS 1388
DB 1642 SKRGSPSPGSSSTESSPEHPKSRFARGRSSSPKTKSRTPRRRS 1690
RESULT 10
Q9JLT1 PRELIMINARY: PRT; 4880 AA.
AC Q9JLT1;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MULTIDOMAIN PRESTYAPIC CYTOMATRIX PROTEIN PICCOLO.
OS Rattus norvegicus (rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20170257; PubMed=10707984;
RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT "Piccolo, a presynaptic zinc finger protein structurally related to
RT Bassoon".
RL Neuron 25:203-214(2000).
DR EMBL; AF138789; AAF07822.2; .
DR HSSP; P04410; 1A25.
DR InterPro; IPR000008; C2.
DR InterPro; IPR002965; PDZ.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00168; C2; 1.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR SMART; SM00239; C2; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00499; C2_DOMAIN_1; UNKNOWN_1.
DR PROSITE; PS50004; C2_DOMAIN_2; 1.
DR PROSITE; PS50106; PDZ; 1.
KW Matrix protein.
SQ SEQUENCE 4880 AA; 530140 MW; 73951E4ED83EA68 CRC64;

Query Match 5.4%; Score 396; DB 11; Length 4880;
Best Local Similarity 20.4%; Pred. No. 3.9e-12;
Matches 323; Conservative 199; Mismatches 563; Indels 502; Gaps 73;

QY 23 PTIDAAYAIKKKKKPPFLPEERSSSEEDDPTPAELLCLICDINTMDAV-TPCCGN-S 80
DB 490 PTKSISQIVTGRPLDP-----PTSAQPPAQLSKTICPLCN--TTELLHPIPERANFN 542
QY 81 SCDECIIRTT-----LLESDKHTCPTCHQN-----DVSP 108
DB 543 TCEGQSTYCSLCGFNPNHILTEIKEMILCLINCMQRAIGDGLAALIPSSQPTPKAATAP 602

QY 109 DALIANKFLROAVNNFKNE-----TGTYTKRLKQLPPFLVPPPPRLSORNQPRSRPI 164
DB 603 TATASKSPVPSSQASPKKPPKODSPKALSKKPP-----EKKKPEPKPPKPPPL 658
QY 165 LRQODPVVRYRTVSPICSGSTAGSCSDGTLRLPAPISLTSNQSILAPVSGN--- 221
DB 659 VQO-----PILHG-PITPATAPQLVAALPEP-----APPKESGGL 694
QY 222 PSSAPAPVDI-----TATVISVHSEKSDGP-----FRSDNKLPAALATSEHSCGA 270
DB 695 PROAKAPVDVBPQKQKMETRADIQSSSTTKPDILSSQVQAOQKATSLKTDASKPS 754
QY 271 SSIIATLALMEKGVCTSPWNSTIFVQSLHQLIPTTPPVIRINARPGGGRGWHSNK 330
DB 755 QSPPTPG---EK---TTPLDISKAMPRPASDskii-----SQPTPG--- 788
QY 331 LGVLVSPQOIRRGERSCYRSINGRHNSERSQRTOSPSPATPCFVPVPPPLYPPPH 390
DB 789 -----SESKDPKHIDPIQKDKPK-----KAQPKSGPKPE 818
QY 391 TPLPPEGVPPQFPSSQPTAGYVPPGFPAPANISTA-CFSPGVPTAHSNTMP 449
DB 819 TKVPKGSPT-----PSGTRPTAGQAAPPQQPKPQEOSRFRSLNLGJITDAPKQSP 871
QY 450 TQOAPLISREEF-----YREOND-----KGRESKFPYSG-SSYSRSSTYDS-----SQGLA 494
DB 872 TTPQETVTKLFGFASISQASNLISTAGQOGPHQTPAPASKQAPPSQAPPAQGRA 931
QY 495 OIHIALTLSPSAANTDLHDHPPEEAARSAMIVHMDLMDIAHARSPPRYRRYS 554
DB 932 KSIQGLPPAPAKATAV-----KKAQKAAA-----E 957
QY 555 RSKSPPEFRGQSPTKRNVPREKEREYFNKRYREVPPIYIKAYGSRVDFRDFEKERYR 614
DB 958 NLESKPE---QAPTAKTTEKDKK-----PPP---AKVGKP---PPSEBEPK-- 993
QY 615 EMERKRYEMVEYKGYAAGA-----QPRPS----- 640
DB 994 -----AVAHKPKDQTKTPKPKPCPLCTRELNLGSGPEPPNFTCTECKNO 1036
QY 641 -ANREDFSP-----ERLLPLIRNSPPTGRGREDYAAQSHRNRLGNYPEKILSTRD 692
DB 1037 VCNLCGFNPTPHLETIQEWLCLNCQORALISGLGM-----GKNPPAS--- 1081
QY 693 SHNAKDNPKSKEKSENVP--GDGKGNKHKHARRRNEKEGESSEF---LNPELLTSR 747
DB 1082 GPKASPMPAPAPBPSQKPTGTQVKGKKKEAEQTEAEKVPVEKETASIEKTPPMVTTDO 1141
QY 748 KCRGSGCIDETKTDTLFLVLP SRDQATPVADPEPDASTIFKYSV-DKDKRE---KDKPKV 803
DB 1142 KLEESG---KSKYSALPEK-----KSEEE---KALSADKKRKPPEAEKPP 1185
QY 804 KSDK-----TKRSDGSATAKKDNVLYK-----PSKGQPKV---GQDRKRSRSP 846
DB 1186 EEKKPLPVQKKLPPEAKPLPSSBGEKHEILKAHVQIPEEPPIGKVAAGKEEQQDSRP 1245
QY 847 PLKKAKEATKIDSVPRSSSQKDEKVTGTPRKRAKSAKADPRROSOPTRRSKRTVPRT 906
DB 1246 ---EALPGATPLLPK---AGEKERA VAPQAGSSKDGGER--SKETEKEDKSDS 1297
QY 907 SSQKSQPVTRPRPSLRKINYL---IAREKNEEKKKSKYDKDFESSMKISVGEETEV 963
DB 1298 SSG--QP---KSPQGLSDTGYSSDGISGLPSLIPSEKDLGLKLDKDSFSQSSPS 1352
QY 964 KPEPKRKMEDV-----EKLETPPEKDIASSTPPAKKIKLINRETKKIGNA 1010
DB 1353 SPEDLAKLSTVLSIIEAOASTLVGEKAEEKTPOKI---SPEK-----P 1394
QY 1011 ENASTTRPESEKLESTSSKIKQEKVKGAKRKVAGSESSSTLVDTSTSTGSGSPVRRS 1070
DB 1395 QDOQKTQTASETLIDIT---ISSEIEIKESQEKKYSPPKDSQ-----GFPSSRKE 1439

Qy	1071	-EETD-----	-KRRVYIKMEEYNNDNTPAPEDVIMIQVQSKWKDKDDESEEDPKT	1123
Db	1440	HKEKPELVLDLSPRRASYSDEVSEDSSENSPVVRRKRTSTIGS--	SSDEYKQEDS----	1493
Qy	1124	TOPLOSQKSPSSITKNWTTTPSATAKYTEKESQEPKLOKLPKEAS--	HELMQHELRSK	1181
Db	1494	-----QSGGEEDDFIRKOLIIEMSDADNASGDE--	EFITSQKLETSGGEGSQKREAKGK	1547
Qy	1182	GSASSEKRAKRDHREHSGSEKDNPKRKSQAQDPKESTVDRLSEQGHFTLSQSKKE----	1237	
Db	1548	GKGVAGCHRRRLTKRSTSPFDDAGRRHSWHDDE--	TFDESELPKFRFTKSQSEPELVYA	1606
Qy	1238	-----	TRISEKHESKSGSN	1252
Db	1607	GGGGLRRFKTELTNSTIADKYSSSSQKTTIYFDEPELMESSLTDSPEDRSRQEGSS	1666	
Qy	1253	---KDFTPGRD---KKVDYDSRDYSSSKRDBRGELARKDPPRGKESU-----	SGQ	1299
Db	1667	LHASFPTPGISPTVSLSLDDSDSSPSHK---KGESKQKRAARRSHGPLPTEDSSE	1722	
Qy	1300	KSRIKREERDLPK-----KGAESKSKSSPPRPDKPHD-----	HKAPYETKRCPEETKPYDK	1350
Db	1723	EEELREBEELIKEGEKRELEQQQRRSSKSKSKKDKELARRRERKTPPSNLTIED	1782	
Qy	1351	NSGKREKHAALRNKRESSGCANCHY	1377	
Db	1783	ASPTTELROAAEM---ELHRRSSCSEY	1806	
RESULT	12			
Q9PU36				
ID	Q9PU36	PRELIMINARY;	PRT:	5120 AA.
AC	Q9PU36;			
DT	01-MAY-2000 (Tremblrel. 13, Created)			
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)			
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)			
DE	ACONIN (FRAGMENT).			
GN	AC2.			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
CC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
PC	TISSUE=BRAIN.			
RX	MEDLINE=99439764; PubMed=10508862;			
RA	Wang X., Kilschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,			
RA	Killmann M.W.,			
RT	"Azoxin, a 550-kd putative scaffolding protein of presynaptic active			
RT	zones, shares homology regions with rim and bassoon and binds			
RT	profilin."			
RL	J. Cell Biol. 147:151-162(1999).			
DR	EMBL, Y19187; CAB60725.1; .			
DR	HSSP; P04410; IAA5.			
DR	InterPro: IPR000008; C2.			
DR	InterPro: IPR001478; PDZ.			
DR	InterPro: IPR002965; P_rich_extensn.			
DR	InterPro: IPR001565; Synaptotagmin.			
DR	Pfam; PF00168; C2; 2.			
DR	Pfam; PF00595; PDZ; 1.			
DR	PRINTS; PRO0360; C2DOMAIN.			
DR	PRINTS; PRO1217; PRICHEXTENS.			
DR	PRINTS; PRO0399; SYNAPTOTAGMIN.			
DR	SMART; SM00239; C2; 2.			
DR	SMART; SM00228; PDZ; 1.			
DR	PROSITE; PS00499; C2_DOMAIN_1; UNKNOWN_1.			
DR	PROSITE; PS50004; C2_DOMAIN_2; 2.			
DR	PROSITE; PS50106; PDZ; 1.			
FT	NON_TER	1		
SEQUENCE	5120 AA; 560751 MW; A65BD9891B65B412 CRC64;			


```
Db 1290 QASLEA-----VEVPSMASSWCGPHFSPEKHEL-SNSPLRENSFGSP---LEFRNSGPL 1339
Qy 1146 ATAYTEKESQPEK-----LOKLPKEASHELMHOLRRSSGASGSKG----- 1189
Db 1340 GTEKNTGSSSEVKEDLNGPFLNQLJETDPSLDMKEOSTSS-GHSSSELSPDAVEKAGNSS 1338
Qy 1190 -----RAKDRHSGSE-----KD---NPKDR-KSGAOP---DKESTVDRLS 1223
Db 1399 NQSTSSPVLAVPRTPSRERSASSAPEMKDGFLRTPSRRSRSGSSPGLRSGSTPSRHS 1458
Qy 1224 EQGFH-----KTLSSSKTRTSEKHEVRS-----SNKDTFGRDKV 1263
Db 1459 LSGSSPGMKDIPRTPSRSECDSSPEPKALPQTPRPSRSGSSPELNKCLTPOQERSG 1518
Qy 1264 DYDSRDY-----SSSKRRD-----ERGEL-----ARKSDPPRGKESL 1296
Db 1519 SESSVDQKYVARTPLGORSRSGSSQELDVKPSASPOERSESDSPDKAKTRTLRQKSR 1578
Qy 1297 SG-----QSKSLR-----EERDLPKGAESKSSNSPPRDKKPHDKAPYETKR 1340
Db 1579 SGSSPEVDSKSRSLRPSRSSSGSSPEVADKRAAPRAQSGSDSSPEPKAPAPALRPSRS 1638
Qy 1341 PCEB-----TKPVDNKSGKERKHAAPRNGKESGANCHVYLTROTLPRRS 1388
Db 1639 GSSSKRGSPSESGSSSTESSPEHPKSRRTARSGRSSPEPKTKSRTPRRRS 1690

RESULT 15
Q95Y2M2
ID Q95Y2M2 PRELIMINARY; PRT; 17352 AA.
AC 095Y2M2
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE I-CONNECTIN.
GN I-CON.
OS Procamburus clarkii (Red swamp crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacidea; Cambaridae; Procamburus.
OC NCBI_taxid=6728;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=21423462; Pubmed=11532946;
RA Fukuzawa A., Shimamura J., Takemori S., Kanzawa N., Yamaguchi M.,
RA Sun P., Matsuyama K., Kimura S.;
RT "Invertebrate connectin spans as much as 3.5 micrometer in the giant
RT sarcomeres of crayfish claw muscle.";
RL EMBL: AB055861; BAB64297.1; -.
DR EMBL: 17352 AA; 1962348 MW; 4BA157BEC042E42D CRC64;
SQ
SEQUENCE 17352 AA; 1962348 MW; 4BA157BEC042E42D CRC64;

Query Match 5.14; Score 375; DB 5; Length 17352;
Best Local Similarity 18.94; Pred. No. 2.2e-10;
Matches 291; Conservative 180; Mismatches 602; Indels 466; Gaps 56;

Qy 19 KYAIPITDAEYAIAGKKEKPPFLDE-----EPSSSEEDDIPPAELL 60
Db 5880 EFSIPKV-----ALKTKVPKKFVPEEKEKLESVDLEHVEKPKAPPTKEKREWSMP----- 5930
Qy 61 CLICDLMTDAVAVIPCGNSSCDCCITRTILLESKHPCPTCHONDVSPDALIAKFLRQA 120
Db 5931 -----DETAYAP-----DELPEREVELEKPP--DTFEPTERKKDEKEOKDKYOR 5973
Qy 121 VNNKFNKGTYKRLKQDLPPFLFLVPPRPLSQNLQPRSRSPILRQODPVVFERYTSP 180
Db 5974 KPRKKTGEYGEKKRLKLGKGR-----PPEEEGEQIKLKPGKRLPQGPBEE----- 6018
Qy 181 CSDTKTAGSCSDGTLRLPAPSISSLTSMOSSLAPVSGNPSSAPAPVDPITATVSI 240
Db 6019 -----KEAPHLKPIPRRKRPEDEKVKPLK-----GKRPSKDIPEREVEVL 6060
```

```
Qy 241 HSEKSDGPFRRDSNKLPLPAALTSSEHSGASSIATIMEKGVGTSPWNISJFVGSL 300
Db 6061 E-----PFEVSDSLIDKQVQLE-----EPVAPBEKLEK--PSTPEGPRI----- 6100
Qy 301 HGOLIPPTGVRINAARPGGRCGWEHNSKLGVLVSPQOIRREKRCYCSINGRHSE 360
Db 6101 -----RWEEKLEEBEGKPRPKKERT-----KPEKEESELPSW-----RGRLLPP 6140
Qy 361 RSQRTQSPSLPAPTCFVVPVPPPLYPPEPHY--LPLPGVPPOFSQPPSOPPTAGYS 418
Db 6141 KEEKEVQOL--KPEKPKPLKP--KPKPKHKPGPIPEPELPEPKTLEPLDKKMQ--- 6194
Qy 419 VPPGFPAPANISTACFSQVPTAHNSMTQTQAPLLSREEFYREGD----- 467
Db 6195 -ITPDHBPQ-----VEDEDEKLPSDBQPLKKKAAVAPKDEEBEIPSMGRK 6242
Qy 468 -----KGR-----SKFYSGSSYSRSYTD 488
Db 6243 LPPEDEKEEITLLKPFKKVVPKPEKPSPLKPGKPEPEIPEPEKSPLEYSKPEKESD 6302
Qy 489 SSQGLAQHIALTLSPSAHTLDLHDHPHPEEAARSAMIVMPDLMDIAHRSRSP 548
Db 6303 KIPELAKVSKRPEPEKPEPEPEKLDSDKKPDESEI--ITQK----- 6344
Qy 549 YRRYRSRSRSPPEFRGOSP-----TKRNVPEEKEREYFNRYRVP----- 589
Db 6345 -KRLQKLKAPTEKEFELPKVTLAKTSQKVFVPEVYLVTELEHVYTPPEVEVEKRV 6403
Qy 590 -PPYDIKAYYGRSVDFRDPPEKERYREWEKRYREWEYKYYKGAQAQPPSANREDFS 647
Db 6404 WSPPEYETVYVEEIPKEPEVELEKYE----- 6433
Qy 648 PERLLPLINSPPTRRRREDYVAGOSHRRNRLGNYPEKLTSDSHANAKNPKESKES 707
Db 6434 -----PPTKPKDEE-----EEDKGYERK-----PKDPEPEDRK 6464
Qy 708 ENYVGDKG-----NKHKKHRRRNEKEGESESFNPELLETSRKRGSSGIDET 758
Db 6465 LKL---GKRLKRPBEEBEEKLKRPKRPRPSKEAE--KPOLKPLPK-----KPEEE 6513
Qy 759 KTDTLFVLPSSRDQATPYVD-EPMDAESI--TFKSVSKDKREKDKPVKSDKTKRKSDG 815
Db 6514 KKDQVTPKPGKPKSKIPDEBPVLEPPEPTEPILDKDVLEPKLPKPPPEPEKPSIE 6573
Qy 816 ATAKKQVLPKSKPOQKVGODREKSPSE-----PLTKAKEEA--KIDS 860
Db 6574 PEAKPLPEKPSSEEBEKLKKEKPKPEKEBEAEVPSWGRKRLPPEKEKEEYLAKPK 6633
Qy 861 VKPSSSQDKDEKVTGTPRKAHSAKDJTRQSOPTRTKRSKTYPKTSSQKSOQVYTRRPR 920
Db 6634 EKPEEPKPKYKKGKGYEPIPEPEKT--PLEEYTPDKKEKVDGTYEPVKKPEDEKPK 6691
Qy 921 SLRKI---NTLIAREKNE-----REKRRKSVDKDESSSMKISKYGEIYKPSPKR 970
Db 6692 PEEIIRKPKRIRKDKDEEVEETPSWRGKRLPPEKEDEEITLLKPFKKEKKEPEPSPIK 6751
Qy 971 -----MEGOVEKLE--RTPEKDKIASSTPAKKIKILNREKGIKGAENASTTKEP 1019
Db 6752 PGKPEPVIEPEKPPLEPEKPEREKVPPQVAPKPEKAPVEO-----PEPEKI 6802
Qy 1020 SEKLESTSSKIKOKVYKGAKRKAVAGSESSSTLVDTYSTSGSGVPRKSEKTDTRKT 1079
Db 6803 SPKPEEEKPOOLEIKIGKRRK-----PRMLAAT--EKFEIPLKI 6842
Qy 1080 VIKTMEEYNDNTAPADVITIMIQVPSKWDKDFESEEDVKTQPIQSVGKPSIKN 1139
Db 6843 TLKK-----TQNVFVPEVYLE-----TIELHVEYVEI-----PEVEKR 6880
Qy 1140 VTKRPSATAYT-----EKSEQPEKLOKL-----PRAASHLQHHLRSKGSASSE-- 1187
Db 6881 WSPPEYETVYVEEIPKEPEVELEKYEKYEPTTKPDEEBEEDKGYERKPKDKPEPED 6940
```

```

Oy 1188 -----KGRKADRHSSSEKDNDRKSGAOPDKSTYDRUSEQHFTLLSSSKNERTS 1241
Db 6941 RKLKLGKGLKPRBEEGEKKLAPKPRK-kspeke-----AEKOLYPIPKKPRBEEK 6993

Oy 1242 EKHSVSGSSNKDFTPEPRDKVUDYSRDYSSSKRRD-ERGELARRKSDPPRKSELS-- 1297
Db 6994 DKVTPRKGGKKRPSKKIIPDR-PELEPEFERTPEPILDKOVPLEKTLAKPKPEKKEPSTER 7052

Oy 1298 -----GOKSKLREERLPPKKABD-----KKS 1319
Db 7053 EAPKPLEEKPESEBEPKLLKERRKPRKPEKEEBEPVPSWRGKRLPRPKBEEKKEIYLKPFKE 7112

Oy 1320 NSSPRDKKPRHDKAPRYETKRRPCEEFKRPVDDKNGSKREK 1358
Db 7113 KPEBPKRKYKPEKRYEPERIPPEKTPLEPYTKPKPKK 7151

```

```
Search completed: September 12, 2002, 19:12:09
Job time: 10377 sec
```

THIS PAGE BLANK (USPTO)